Volume 34 JULY, 1939 Number 3

The Journal of Abnormal and Social Psychology

EDITED BY

GORDON W. ALLPORT
HARVARD UNIVERSITY

CONSULTING EDITORS

EDNA HEIDBREDER
Wellesley College

FLOYD H. ALLPORT
Syracuse University

RUTH BENEDICT
Columbia University

HENRY E. GARRETT
Columbia University

JAMES Q. HOLSOPPLE
New Jersey State Hospital

CARNEY LANDIS
New York Psychiatric Institute

HENRY T. MOORE
Skidmore College

DAVID SLIGHT
University of Chicago

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITORIAL OFFICE: EMERSON HALL, CAMBRIDGE, MASS.

PUBLICATION OFFICES: ALBANY, N. Y., AND
THE OHIO STATE UNIVERSITY, COLUMBUS, OHIO

Entered as second-class matter, June 18, 1923, at the Post Office at Albany, New York

FACTORS DETERMINING SUBSTITUTE BEHAVIOR AND THE OVERT EXPRESSION OF AGGRESSION

BY LEONARD W. DOOB AND ROBERT R. SEARS
Institute of Human Relations, Yale University

THE proposition has been presented elsewhere (2) that aggression is a response to frustration. According to this view a frustration is any interference with a goal-response, whether the interference be caused by some hampering agent or by a conflict of responses within the organism itself. The frustration produces unique changes within the organism which serve as instigation to aggressive behavior. The strength of the frustration-induced instigation, in turn, is reduced by this aggressive behavior and may be called *secondary* to distinguish it more readily from the *primary* instigation to the original (frustrated) goal-response. The situation is similar to that of any response which is dependent on a specifiable drive. When a child, for instance, is hungry, it may be said to be instigated to eat; after carrying out the act of eating, the instigation is presumed to have been reduced in strength, since the eating behavior stops. Aggression differs from many other goal-responses, however, in that the instigation to it is produced by interference with any other pre-established goal-response rather than by recurrent intraorganismic changes such as those involved in hunger, thirst, demands for social approval, etc.

The importance of this distinction between primary and secondary (frustration-induced) instigation lies in the systematic differentiation of substitute responses from aggression. That all frustrating situations do not produce overt aggression is immediately evident. An employee who fails to receive an expected promotion is more likely to double his efforts or search for a new job or treat himself to some small luxury than to berate his employer in a face-to-face conversation. While the latter response would be typically aggressive, the more frequently occurring responses to such a frustration seem to contain little or no aggression and may be more appropriately classified as substitute

responses. They are substitutes in the sense that they tend to reduce the strength of the primary instigation to the original frustrated goal-response. Redoubled efforts and the search for a new job permit the development of new anticipations of promotion, anticipations which in themselves are rewarding, i.e., constitute goal-responses. The indulgence in a small luxury is a fractional part of the total set of goal-responses which would have occurred had the original frustration of non-promotion not occurred. Such responses may be facilitated by the secondary, frustration-induced instigation but they are, in the main, responses to the original, primary instigation to economic self-betterment.

In so far as any substitute response is "pure," i.e., contains no aggressive component, it will reduce the strength of the primary instigation. Aggression reduces the strength of the secondary, frustration-induced instigation only. But since substitute responses reduce the primary instigation, they likewise reduce the strength of the frustration (2, Chap. I) and hence reduce the probability of occurrence of aggressive behavior.

In many frustrating situations, however, it is difficult to find evidence of either substitute responses or overt aggression in the behavior of the frustrated person. It is necessary, consequently, if the frustration-aggression hypothesis is to be maintained, to postulate some form of unexpressed or *non-overt* aggression. And, in fact, brief questioning of a frustrated individual will often elicit such statements as "I'm angry" or "I'd like to get even with him" or "I'll kill him in my dreams for a week!" While these do not involve any immediate aggressive action against the frustrating agent, they must nevertheless be considered as aggressions from the conceptual standpoint.

It may be said, then, that frustration is followed not only by *aggression which may vary in the degree to which it is expressed overtly,* but also by *substitute responses* as well. The problem attacked in the present study is the isolation of factors which determine the occurrence of these possible alternatives as reactions to frustration and, in the case of aggression, the factors which determine the occurrence of overt or non-overt aggression.

Two hypotheses may be suggested to account for the occurrence of substitute responses rather than aggression: first, that the

frequency of substitute responses varies positively with the strength of anticipatory responses to punishment-for-being-aggressive; and second, that the frequency of substitute responses varies inversely with the strength of the frustrated goal-responses.

First hypothesis. It is clear that almost any form of behavior can be suppressed or eliminated by punishment; the law of effect is the most familiar statement of this fact. During the process of learning, an organism develops anticipatory responses to the punishment which on later occasions serve to block or inhibit the proscribed behavior. *It is these anticipatory responses to punishment that are reflected in verbal judgments of anticipating or expecting.* This first hypothesis implies that aggressive behavior is punished more frequently than substitute behavior. While this contention would be difficult to prove categorically even in our own relatively well-documented society, certain logical necessities give the assumption an appearance of validity. Substitute responses are by definition actions that are relatively non-aggressive; when other persons are involved in the situation, therefore, they are not so frustrated as they would be by aggressive actions. Since they are less frustrated, they, in turn, are not so likely to become counter-aggressive. As a consequence, relatively less counter-aggression, or punishment, is anticipated by a frustrated individual when he responds to the frustration with a substitute response than when he responds with overt aggression. If, then, substitute responses are less often followed by punishment [1] and if the amount of punishment received in the past for a given act determines the extent to which it is eliminated, there should be a greater frequency of substitute responses when subjects express a greater amount of anticipation of punishment for being aggressive.

Second hypothesis. If the above principle is true, it follows that an increase in the strength of instigation to the aggressive responses could overcome the inhibiting strength of the anticipatory responses to punishment. Hovland and Sears (3) have suggested that the strengths of such incompatible or antagonistic responses summate algebraically; and, since in this case the strength of the aggressive responses varies positively with the

[1] All substitute responses are not socially acceptable (e.g., some forms of delinquency, compensations for felt inferiority, etc.), but in general substitutive behavior appears to interrupt the activities of others less than does aggressive behavior.

strength of the frustrated goal-responses, it can be concluded that the amount or frequency of substitute responses will vary inversely with the strength of the drive.[2]

The factors determining the overtness of aggression are somewhat similar to those determining the occurrence of substitute responses and may be expressed in the form of two further hypotheses: third, that overtness varies inversely with the amount of punishment anticipated as a consequence of such behavior; and fourth, that overtness varies positively with the strength of the instigation to the frustrated goal-response.

Third hypothesis. The reasoning which leads to the formulation of the third hypothesis is essentially the same as that described for the first, except for the fact that the assumption concerning the kind of behavior most often punished relates to overt and non-overt aggression rather than to aggression as a whole and substitute responses. Evidence such as the following lends support to the belief that overt aggression is frequently punished in American society. Wickman (6, 113, 127) obtained 511 teachers' ratings on the relative seriousness of a list of 50 school behavior problems. Each problem was rated on a 20-point graphic-rating scale. Of the 50 behavior problems, 12 were clearly overt forms of aggression and were given an average rating of 12.3. The remaining 38 problems (including three which might indicate non-overt aggression: "unhappy or depressed," "easily discouraged," "sensitiveness") were given an average rating of 11.1. Overtly aggressive behavior was placed, therefore, in the upper half of the list in terms of seriousness. That this finding probably reflects the teachers' attitudes toward the reprehensibility of such behavior is indicated by the fact that 30 child psychiatrists judged the same items as having a seriousness rating of only 8.9, while the remaining items were given an average rating of 9.7. If these 511 teachers from New Jersey, New York, Ohio and Minnesota can be considered a fair sample of the disciplining group in the American classroom, it may be said with considerable certainty that overt aggression meets with serious objection from those who are in a position

[2] In a previous communication (2) the postulate that *the strength of any frustration-reaction, including aggression, varies positively with the strength of the instigation to the frustrated goal-responses* was presented as necessarily fundamental to any quantitative theory of conflict and frustration.

to punish.[3] Direct evidence that people anticipate more punishment for overt than for non-overt aggression has been obtained from the present study and will be presented in the second section of the Results. From this reasoning the hypothesis is derived that the amount or frequency of overt aggression varies inversely with the strength of the anticipatory responses to punishment.

Fourth hypothesis. Here again reference must be made to the assumption that the strengths of incompatible responses summate algebraically. When the strength of the frustrated goal-responses and hence that of the overt aggressive reactions is great, the inhibiting influence of the anticipatory responses to punishment may be overcome, with the result that the overt aggression will occur. The greater the strength of the original goal-responses, the greater is the probability of this outcome.

Although the terms overt and non-overt aggression have been used freely in describing the particular kinds of behavior here studied, it must be remembered that they refer to extremes of a continuum. In actual fact it must be presumed that there are very many discriminable divisions on the scale of "degree of overtness"; but until some method for equating and ordering these divisions is discovered, only the extremes may be safely used. In the discussion below the following abbreviations are used:

O — Overt aggression
N — Non-overt aggression
S — Substitute responses

The variables presented in the four hypotheses above by no means exhaust the possible list of factors which determine the kind of response to be made in a frustrating situation; they represent simply first approximations. Most seriously lacking is an hypothesis to account for the occurrence of substitute responses rather than non-overt aggression when overt aggression has been punished. Since, however, this differentiation probably is dependent on complex learning experiences during long periods of development, the particular technique of the present investigation has not seemed applicable to a study of the problem.

[3] Some forms of overt aggression are permissible and encouraged in our culture (e.g., competitive sports, school spirit, chivalric aggression, etc.).

Additional factors of possible importance in determining overtness of aggression may be suggested.

1. Degree of anticipation of success in overcoming the frustration by being overtly aggressive.
2. Degree of submissiveness ("invitation to aggression") on the part of the frustrating agent.
3. Frequency of the given frustration in the subject's past experience.
4. Degree to which generalization of response exists in the subject; i.e., the degree to which he has a *trait* of being overtly or non-overtly aggressive.

These suggested factors fall within the systematic structure of a reaction psychology and hence are congruent with the four hypotheses tested in the present study.

It has been assumed elsewhere (2, Chap. 3) and it is assumed here that overt, direct aggression against the agent which is perceived as the source of frustration is most satisfying to the organism, i.e., is the response to which there is strongest instigation. Whether this tendency to avenge a frustration by attacking the frustrator is original or learned by people in various societies is a question which must be raised but which only future research can answer.

It is known, however, that the most satisfying response is not always the one which is tolerated by society; socialization, in fact, has as one of its functions the inculcation of taboos—the anticipation of punishment for carrying out certain acts (2, Chap. 4). If, even though individuals find a mode of behavior very satisfying, its occurrence is simultaneously inhibited, then it is reasonable to conclude that they have been socialized to the extent that they are able to anticipate punishment from acts which they also anticipate to be satisfying. Evidence concerning the degree of satisfaction and amount of punishment to be anticipated as a consequence of overt and non-overt aggression and substitute responses is given in the second section of the Results.

A psychological systematization of the problem, however, is not necessarily the unique method of formulating it. From a sociological standpoint the question might be phrased: to what degree are various situations frustrating and to what degree

do they lead to substitutive behavior or to the overt expression of aggression? In this case, where the *situation* as a measurable social phenomenon is made the key to the problem, an examination can be made either of the behavior of representative members of a society in terms of their conformity to cultural norms or of a situation in the society itself in terms of such behavioral variables as the overtness of aggression. The latter type of analysis permits generalizations concerning culturally described situations; these generalizations not only have behavioral relevance but actually explain social and cultural phenomena by reference to the laws of behavior of the individuals who compose the social body portraying the phenomena. An example of this method of interpretation of social psychological data is given below in the third sub-section of the Results.

METHOD

To test the four hypotheses presented above, data were obtained concerning the previous behavior of subjects in certain frustrating situations which had been fairly common in the lives of all the members of the group. To facilitate uniformity both in the subjects' understanding of the described situations and in their techniques of recall and estimation, a standardized form was prepared on which each subject could record the data from his own recollections.

Construction of the questionnaire. A brief description was prepared of each of 53 situations which either on *a priori* grounds or because of their presence in Cason's list of annoyances (1) seemed to be definitely frustrating and to occur fairly frequently to Yale College students. Three of these descriptions exemplify the list adequately.

1. A policeman reprimands you vigorously and unjustly, perhaps gives you a ticket, for a traffic offense when you are accompanied by a girl whose good opinion you value.
2. You have made extensive and expensive plans for a date. The girl informs you at the last moment that she cannot come and the reason for her breaking the date seems a very inadequate one to you.
3. An instructor announces an unexpected examination for the next meeting of the class.

The list of situations was given to a group of 30 Yale College

students enrolled in an elementary course in Personality. The instructions were as follows:

1. Read each item carefully.
2. After reading an item, stop for a moment and recall the last occasion when that or something very similar happened to you.
3. Describe briefly what you did in that situation, what your thoughts were, and what your emotions (if any) were. If you have never found yourself in such a situation, mark a cross in the blank space beneath the item.

From the answers obtained in this fashion, sixteen test items were selected for use in the final form of the questionnaire. Two criteria dictated their selection: first, high frequency of occurrence as indicated by the number who answered each item; and second, the occurrence in each situation of all three kinds of response, O, N, and S. A large number of items had to be discarded because they evidently occurred so infrequently in the students' lives that even in a large group only a few persons would have suffered them. A lesser number were discarded because they elicited, too uniformly, only overt or non-overt aggression or substitute responses.

From these free descriptions at least one characteristic reaction of each of the three types, O, N, and S, was obtained for each of the sixteen situations; for most of the sixteen situations more than three alternatives were secured, but all were classifiable as one of the three types of response (O, N, S). Each reaction was phrased as briefly as possible (see examples below) to insure that the alternatives should be sufficiently extreme in their respective categories. Five psychologists who were familiar with the descriptive concepts employed were asked to rate each alternative under each situation as O, N, or S. Among these judges there was almost complete agreement. Such difference of opinion as there was appeared to derive largely from uncertainty as to the "purity" of a reaction's classification; for example, several of the S reactions were judged to involve a slight degree of N as well. These ambiguities were largely eliminated by rewording the descriptions of the responses.[4]

[4] Several of the S alternatives appear to represent "substitute gratifications" rather than "substitute responses" to the original instigation. The systematic relationship between these two concepts is not entirely clear at present, and it may be that they operate according to different laws. Since the only effect of combining the two concepts can be to attenuate the differences reported under the first two hypotheses, no attempt will be made to analyze the data differentially in terms of the two concepts.

For each situation a four-point rating scale was prepared on which could be answered quantitatively a question as to the strength of the instigation to the frustrated goal-responses. Each of these questions was examined by the five judges to determine whether it referred accurately to the instigation of the frustrated goal-responses.

Final form of the questionnaire. The page of instructions used with the final form of the questionnaire included a sample situation which was not scored in the tabulation of results. The form of presentation of each of the other sixteen situations was the same as that of the first three, which are shown below to indicate more adequately the types of responses presented as alternatives for the subjects to check. After each response is shown in parentheses the category to which the response belongs.

1. You saw an item in the paper which described a case of conspicuous and unnecessary brutality and cruelty on the part of the police.
   - I boiled up inside and wanted to do something about it but did not know what I could do. (N)
   - I thought what I would like to do to the police. (N)
   - I wrote a sharp letter to the newspaper, condemning such actions. (O)
   - I quickly began to think of something more pleasant. (S)
   - I protested to the chief of police. (O)
   - I felt angry. (N)
   - I talked to my friends about the brutality of the police. (O)

   How strongly did you feel that the police should not use violence?
2. In a semi-formal group you heard a sneering comment about Yale which you felt was unjustified. You discovered that the speaker was a stranger whose appearance you disliked anyway.
   - I made a slurring remark in return to put him in his place. (O)
   - I unobtrusively left the group. (S)
   - I was disgusted or felt resentful, but did not go to the trouble of challenging him. (N)
   - I ignored him and diplomatically turned the conversation to another topic. (S)

   How warmly did you feel toward Yale at that time?
3. Someone sitting in a seat near you in a theater or concert wiggled and squirmed about in a totally unnecessary way.
   - I made some uncomplimentary remarks about him to my companions. (O)
   - I moved to another seat. (S)
   - He made me sore and disgusted; it almost ruined the performance for me. (N)
   - I told him to keep quiet. (O)

   How much did you want to enjoy the performance in peace and quiet?

The remaining 13 situations were as follows:

4. You were behaving decorously at a party and an acquaintance who was too drunk spilled some food, his drink, or cigarette ashes all over you.
5. You had to get out of bed after you had gone to sleep in order to answer the telephone. It proved to be someone who had gotten the wrong number.
6. You tried to have a calm and rational discussion with a good friend about some rather emotional topic (e.g., religion, politics, morals), but he lost his temper and refused to be calm and rational.
7. A guest of yours whom you knew only slightly persisted in telling you how to drive and, in general, pestered you with "back-seat" criticisms.
8. A good friend of yours was driving so recklessly that you were genuinely scared out of your wits.
9. You were in the midst of a task that was taxing your patience (e.g., changing a tire) and a stranger, trying to be helpful, added his cheerful but utterly impractical suggestions.
10. You were bored stiff by a conversation of friends, but the social situation was such that you had to endure it for several minutes at least.
11. In an informal social gathering of people with whom you were not very well acquainted you found you couldn't solve what everyone else considered a very simple parlor trick.
12. A poorly informed foreigner made disparaging remarks about the United States.
13. A very casual acquaintance near you cheated on an examination.
14. You found that although a friend of yours had written a poorer examination than you, he had received a higher grade.
15. When you finally managed to dance with a particularly attractive girl, a good friend of yours cut in.
16. You desired to board a boat or train to see a friend off, or to enter an exhibition or park; the guard forbade you on what seemed to be entirely unnecessary technicalities.

Instructions. The following instructions were memorized and spoken to an elementary class in social psychology composed of 185 Yale undergraduates:

Today I am asking you to fill out a questionnaire. Please be as honest and as truthful as possible. The results, as I have explained to you, are for research purposes. Your name will be kept secret. Please do not look at the rest of the questionnaire until you understand the instructions completely. You will be given an opportunity to ask any questions about points not clear to you.

Let us use the situation on the first page by way of illustration. You must decide whether the situation has actually occurred to you and whether you remember your behavior in the situation sufficiently vividly. If the situation has not occurred to you or if you do not remember the details of your behavior, skip the particular question. There is no social disgrace, of course, in skipping a question, but please make an effort to answer as many questions as possible.

I am asking you to do four things for each question in which the situation has occurred to you:

1. In the first column [three blank columns were ruled in the left margin] check the item or items which best described or describe your behavior at that time. You check those items which describe your behavior and as many or as few as you wish. If you feel that the alternatives described do not cover your behavior, write what you did in the blank space provided for that purpose.

2. In the second column rate the reaction which would have brought you the most satisfaction if you had done it. Perhaps you actually did what brought you most satisfaction. In any case put your rating next to the item which might have brought you, or actually did bring you, the most satisfaction. Put a rating of 1 if this brought you no satisfaction, 2 if it brought slight satisfaction, and so on. These ratings are listed on the first page and are also here on the blackboard.

3. In the third column, rate the reaction which would have brought you the most trouble (or have caused you the most unpleasant consequences) if you had done it. Perhaps you actually did what brought you the most trouble (or caused you the most unpleasant consequences). In any case put your rating next to the item which might have brought you or actually did bring you the most trouble (or caused you the most unpleasant consequences). Put a rating of 1 if it brought no trouble, and so on.

Be sure to recollect what would have or did bring you the most satisfaction and trouble in columns 2 and 3, as these possibilities occurred to you at the time when the incident took place. Do not reconstruct your behavior on what you think about that incident now. In some cases, it may not be possible to rate columns 2 and 3; try to, but, if you can't, then just don't.

4. At the bottom of each incident you are asked an additional question. Give your answer in the form of a rating by checking the alternative which most nearly corresponds to your feeling at the time. There are four possibilities.

The students were told with emphasis that they should answer an item only if that situation had occurred to them. They were given as much time as they wished to answer; all of them finished within 40 minutes. In scoring the items, those responses which subjects added because none of the standardized alternatives described their behavior were judged independently (as O, N or S) by the writers and included in the final tabulation only if the judgments coincided. Otherwise the situation was not tabulated for that particular subject.

This questionnaire was designed to obtain, in a standardized manner, reports of what had actually happened in real-life frustrating situations. The fact that it was standardized implies at once that the answers were limited in certain ways. The

nature of the hypotheses under examination determined the kind of limitation. In summary, the data obtained from the 185 subjects comprised the following:

1. What the actual response was to each of the sixteen frustrating situations which had been experienced.
2. What, and to what degree, would have been the most satisfying response to the situation, as the subject recalled his impulses at the time of occurrence.
3. What, and to what degree, would have been the response that would have caused the subject the most trouble or punishment as the subject recalled his judgments and anticipations at the time of occurrence.
4. How strong the instigation was to the goal-responses which were frustrated in the described situations.

RESULTS

I

The standardized recall was made by 185 subjects. The average number of situations for which recalls were made was 12.1; the S.D. was 3.2. In Table 1 are given the percentage of subjects

TABLE 1

NUMBER OF SUBJECTS WHO RECALLED EACH SITUATION AND THE PER CENT WHO MADE EACH TYPE OF RESPONSE

SITUATION	N	PERCENTAGE OF SS MAKING EACH TYPE OF RESPONSE						
		O	N	S	ON *	OS *	NS *	ONS *
1	151	20.5	55.0	4.6	18.6		1.3	
2	136	47.8	33.1	13.2	1.5	1.5	2.9	
3	172	50.6	14.5	16.9	8.7	4.7	2.9	1.7
4	128	13.3	30.4	46.1	.8		8.6	.8
5	125	21.6	36.8	32.8		3.2	5.6	
6	151	32.4	19.9	29.8	2.0	.7	12.5	2.7
7	125	72.0	10.4	6.4	5.6	3.2	.8	1.6
8	165	72.1	15.8	3.0	2.4	4.3		2.4
9	123	16.3	38.2	29.3	3.2	2.4	9.8	.8
10	177	5.6	35.0	48.0	.6	1.7	8.5	.6
11	121	9.1	42.2	23.1	4.1	1.7	19.0	.8
12	85	44.7	29.4	21.2	1.2	1.2	2.3	
13	161	13.0	40.3	22.4	5.6	1.9	14.9	1.9
14	152	16.4	51.3	13.8	4.0	4.6	9.9	
15	152	35.5	17.8	23.7	7.9	5.9	4.6	4.6
16	136	16.9	22.1	31.6	5.9	5.1	8.1	10.3
Average	141							

* The subjects represented in these columns recalled having responded to the situation with more than one of the alternatives; when all the reported responses were the same (O, N or S), the responses were scored as a single response and placed in the appropriate category without differentiation from single responses in that category.

making each type of response to each of the sixteen situations. For each situation the percentages are based on the total number of subjects responding to that situation, as shown in the second column of the table.

First hypothesis. The validity of this hypothesis depends on the demonstration that the proportion of substitute responses (S) was greatest when anticipation of punishment for being aggressive was greatest. With the present data this latter variable can be measured only by a comparison of the size of the ratings given in column 2 (anticipation of satisfaction) and 3 (anticipation of punishment) of the questionnaire. Obviously there would have to be some impulse to do a certain act, and an expectation of pleasure from doing it, before the anticipation of punishment could act as a deterrent. To test this hypothesis, therefore, only those instances were used in which the *same* alternative was checked as providing anticipated satisfaction and punishment; i.e., those instances in which the subject both did and did not want to do a certain thing.[5] Since the hypothesis relates to punishment for being aggressive, furthermore, only those instances from this group were used in which the activity both desired and not desired was either O or N; instances which fitted the first criterion but involved S alternatives were ignored, since they were not relevant to the hypothesis. Within this selected group of cases three degrees of anticipation of punishment can be distinguished: (1) that in which anticipated satisfaction was greater than anticipated punishment, (2) that in which the two were of equal strength, and (3) that in which anticipated satisfaction was less than anticipated punishment.

According to the hypothesis, the proportion of substitute responses should increase from the first to the third degree. In Table 2 are shown, for each of these three degrees, the number of instances in which the above criteria were met and the percentage of these instances in which a substitute response was

[5] Although it may be presumed that careful introspection at the time of a frustration would reveal one or several impulses both to do and not to do a certain thing, not all of the subjects reported such conflicts. Some of them checked one alternative in column 2 and another in column 3; it is these subjects who have been excluded in the results pertaining to the first and third hypotheses. It was felt that compelling subjects to rate the same alternative in both columns would have increased the artificiality of the questionnaire. Those who rated different columns were probably recalling inaccurately their mental state during the frustration.

the only response actually made. These same data are given likewise for all the cases under each of the four strengths of frustration.

The uniform rise in proportion of substitute responses from top to bottom in the table is in accord with expectation on the basis of the first hypothesis, although the critical ratios of the differences (in the right-hand column) are small: between S > P and S = P, 1.2; between S = P and S < P, 0.6; between S > P and S < P, 1.2.

TABLE 2

DATA REQUIRED FOR A TEST OF THE FIRST HYPOTHESIS

The rows represent three different amounts of anticipated punishment for being aggressive. The columns represent (N) the number of situations rated as involving each of the four strengths of drive, and (S) the percentage of these situations which evoked substitute responses. The last column is critical for the first hypothesis.

RELATIVE STRENGTHS OF ANTICIPATED SAT. AND PUN.	STRENGTH OF DRIVE-INSTIGATION								TOTAL	
	STRONG 1		2		3		WEAK 4			
	N	S	N	S	N	S	N	S	N	S
Satisf. > Pun.	241	9.5	203	18.2	79	17.7	22	22.7	545	14.5
Satisf. = Pun.	122	9.8	153	19.0	82	23.2	23	26.1	380	17.4
Satisf. < Pun.	26	15.4	37	13.5	13	30.8	8	50.0	84	20.2
Total	389	10.0	393	18.1	174	21.3	53	28.3	1,009	16.0

Second hypothesis. The number of subjects who answered each item with each of the four degrees of drive strength is too small to permit a test of this hypothesis on each of the sixteen situations. In Table 3, therefore, are given, for all the instances rated under each of the four strengths of drive,[6] the percentage of responses which were O, N, S, and combinations of these. The *N*'s in this table represent all the answers to all sixteen situations by all subjects without reference to the ratings made in column 2 and 3 of the questionnaire. The data of Table 3 are the same as those of Table 1. According to this second hypothesis the proportion of substitute responses should increase with a decrease in strength of instigation. That this condition is ful-

[6] The responses of subjects who gave ratings in the "very strongly" division of the scale below each item are given in the column labelled 1 in Table 3; those in the "cared very little" division are given in column 4; and similarly for columns 2 and 3.

filled by the data can be seen from following the row labelled "S" from left to right. The critical ratios of the differences between successive strengths are: between 1 and 2, 3.5; between 2 and 3, 3.3; between 3 and 4, 2.7; between 1 and 4, 7.0.

TABLE 3

FOR THE 185 SUBJECTS, THE TOTAL NUMBER OF SITUATIONS RATED AS HAVING EACH OF THE FOUR STRENGTHS OF DRIVE-INSTIGATION, AND THE PERCENTAGE OF THIS TOTAL NUMBER OF SITUATIONS RESPONDED TO WITH EACH OF THE TYPES OF RESPONSE

TYPE OF RESPONSE	STRENGTH OF DRIVE-INSTIGATION				TOTAL
	STRONG 1 N = 770	2 N = 866	3 N = 444	WEAK 4 N = 180	N = 2260
O1 *	35.8	28.2	24.3	17.2	29.2
O2 *	1.7	1.3		.6	1.1
O3 *	.3	.1			.1
N	28.3	33.1	31.3	26.7	30.6
S	14.7	21.6	30.4	44.4	22.8
ON	6.5	4.2	3.6	2.2	4.7
OS	3.6	2.4	2.3	1.1	2.7
NS	5.7	8.0	7.2	7.2	7.0
ONS	3.4	1.1	.9	.6	1.8

* Multiple Overt responses have been kept separate, in this table, from single Overt responses; O1, O2 and O3 indicate the number of Overt responses checked.

Third hypothesis. Table 4 presents the data in the way required for determining the validity of the third hypothesis—namely, that the overtness of aggression varies inversely with the degree of anticipation of punishment for being overtly aggressive. As with the test of the first hypothesis above, only those instances were considered in which the same alternative was specified as providing the greatest anticipation of both satisfaction and punishment. In this case, again, a second delimitation of instances was necessary; only those instances were used in which the alternative thus rated was an O alternative. This selection is necessary because the third hypothesis relates to *punishment for being overtly aggressive* and not, as does the first hypothesis, to *punishment for any kind of aggression.* The same three degrees of anticipation of punishment were used in the present

test (Table 4) as were used for testing the first hypothesis (Table 2).

The hypothesis requires that the proportion of overtly aggressive responses should decrease with increased anticipation of punishment. In Table 4 are given, for the various degrees of anticipation and of strength of instigation, the number of cases which met the criteria mentioned above and percentages of

TABLE 4

DATA REQUIRED FOR A TEST OF THE THIRD HYPOTHESIS

The rows represent three different amounts of anticipated punishment for being *overtly* aggressive. The columns represent (N) the number of situations rated as involving each of the four strengths of drive, and (O) the percentage of these situations which evoked *overt* aggressive responses. The last column is critical for the third hypothesis.

RELATIVE STRENGTHS OF ANTICIPATED SAT. AND PUN.	STRENGTH OF DRIVE-INSTIGATION								TOTAL	
	STRONG 1		2		3		WEAK 4			
	N	O	N	O	N	O	N	O	N	O
Satisf. > Pun.	214	65.4	164	56.7	61	59.0	14	42.9	453	60.7
Satisf. = Pun.	105	52.4	119	47.1	60	55.0	17	41.2	301	50.2
Satisf. < Pun.	20	50.0	30	30.0	9	22.2	4	0	63	33.3
Total	339	60.5	313	50.5	130	54.6	35	37.1	817	54.7

aggressive responses which were overt (number of O responses divided by number of O + N). Even though the number of instances is small in some categories, the predicted trend holds good, both for every degree of drive strength and for the total, which disregards differences in drive strength. For the differences in the right-hand column of Table 4 (Total) the critical ratios are: S > P and S = P, 2.7; S = P and S < P, 2.6; S > P and S < P, 4.3.

Fourth hypothesis. In Table 5 are given the data relevant to the hypothesis that overt aggression varies positively with strength of drive. As in Table 3, these data represent all the responses of all subjects to all situations. For each of the four strengths of drive are shown the percentage of aggressive responses, the percentage of overtly aggressive responses, and the proportion of overtly aggressive responses to all aggressive responses. Not only do total aggressions increase with increase in drive strength

(the obverse of the first hypothesis),[7] but the proportion of O in relation to N responses increases. The critical ratios of the differences in the bottom line of Table 5 are: between 1 and 2, 5.1; between 2 and 3, 0.8; between 3 and 4, 1.0; between 1 and 4, 4.3.

TABLE 5

IN RELATION TO THE STRENGTH OF DRIVE-INSTIGATION, THE NUMBER (AND PER CENT) OF ALL RESPONSES WHICH WERE AGGRESSIVE AND THE PER CENT OF AGGRESSIVE RESPONSES WHICH WERE OVERT

TYPE OF RESPONSE	STRENGTH OF DRIVE-INSTIGATION							
	STRONG 1		2		3		WEAK 4	
	N	PER CENT	N	PER CENT	N	PER CENT	N	PER CENT
Aggressive	657	85.3	679	78.4	309	69.4	100	57.5
Overt-Ag.	395	51.3	323	37.3	138	31.1	39	21.7
Ov.-Ag./Ag.		60.1		47.6		44.7		39.0

II

It is now possible to determine whether these particular subjects have considered overt, direct aggression to be most satisfying. In 2,134 of the 2,260 situations a single alternative was checked as being the most satisfying response. Of these instances, 69.2 per cent were overt aggressions, 7.4 per cent were non-overt, and 23.4 per cent were substitute responses. Among the alternative responses presented with three (Nos. 1, 3, 13) of the sixteen situations, moreover, were responses which represented both direct and indirect overt aggression, i.e., overt aggression toward the frustrating agent or overt aggression not so directed. A far larger proportion of the subjects reacted with direct aggression than with indirect aggression. In No. 1, the third and fifth alternatives represented direct aggression (88 cases) and the seventh alternative represented an indirect (and perhaps partly non-overt) aggression (9 cases). In No. 3, the first alternative

[7] This principle has been expressed elsewhere (2, Chap. 2) as fundamental to the systematization of the frustration-aggression relationship. It is clear from Table 3 that not only did the percentage of cases in which a single O response occurred increase as drive strength increased, but there were increasing numbers of instances in which two and even three O responses were recorded.

was indirect (13 cases) and the fourth was direct (113 cases). In No. 13, a similar pair of alternatives was available: "I griped about it to mutual friends after the examination" (indirect, 12 cases), and "I told him to stop" (direct, 57 cases).

Not only is it clear from the data that the most satisfying response is the overt, direct form of aggression, but it is also evident that overt aggression gives rise to strong anticipations of punishment. In 2,125 of the 2,260 situations recalled, a single alternative was checked as being the response from which most punishment was anticipated. Of these instances, 85.9 per cent were overt aggressions, 6.8 per cent were non-overt, and 7.3 per cent were substitute responses.

III

It is possible, as was suggested in the introductory section, to analyze the data situation by situation. Obviously the findings have certain limitations: only a small sample of social situations has been selected and the individuals studied represent just one segment of the culture, *viz.*, Yale College students. The following discussion, however, may suggest a method of approach to similar problems. For illustrative purposes, three factors have been isolated. They are presented in Table 6. In column 1 is the percentage of subjects who displayed overt aggression in each of the sixteen situations; multiple responses in which 1 or more of the alternatives were O responses are included. In column 2 is given the percentage of subjects who were "strongly" frustrated in the situation; the criterion for "strong" frustration is a rating of "1" for the strength of instigation to the frustrated goal response. In column 3 is given the percentage of subjects who believed they would secure an equal amount of satisfaction and trouble from doing what they wanted to do; again only those are included who rated the same action in both respects. This third factor can be interpreted in terms of conflict. Since the impulse to do something was as strong as the one not to do the same thing (column 2 = column 3), an internal conflict of Type II variety can be presumed.[8] It cannot be denied that, when the impulse to do something was greater or less than the

[8] For a description of Type II conflicts and their consequences, see Lewin (4, 91–92) and Hovland and Sears (3).

impulse not to do it, a conflict may also have been present; but here the assumption is made that the conflict was at least *stronger* when the impulses had equal strength.[9] The data are given in percentages also and the situations are ranked with respect to each other for each of the three variables. A high percentage means that the situation evoked, respectively, a great amount of

TABLE 6

FOR EACH FRUSTRATING SITUATION, THE PERCENTAGE OF SUBJECTS WHO WERE (A) OVERTLY AGGRESSIVE, (B) STRONGLY FRUSTRATED, AND (C) IN STRONG CONFLICT

The total *N*'s to which the percentages for (A) and (B) refer are given in Table 1; those for (C) are given here.

SITUATION	OVERT AGGRESSION		STRONG FRUSTRATION		STRONG CONFLICT		
	PER CENT	RANK	PER CENT	RANK	N	PER CENT	RANK
1	39.1	7	27.8	10	50	32.0	12
2	50.7	5	44.9	4	95	32.6	11
3	65.7	3	39.5	6	89	27.0	14
4	14.8	15	27.3	11	61	47.5	2
5	24.8	11	26.4	12	57	24.6	15
6	37.7	9	15.2	14	71	39.4	8
7	82.4	1	46.4	3	67	38.8	9
8	81.2	2	70.9	1	65	30.8	13
9	22.8	12	43.9	5	74	46.0	3
10	8.5	16	16.9	13	105	53.3	1
11	15.7	14	14.9	15	30	40.0	7
12	47.1	6	35.3	9	58	44.8	4
13	22.4	13	11.8	16	42	40.5	6
14	25.0	10	48.7	2	93	34.4	10
15	53.9	4	37.5	7.5	55	21.8	16
16	38.2	8	37.5	7.5	76	43.4	5

overt aggression, strong frustration, and strong conflict *as these terms have been defined in this connection.*

In the light of a psychological analysis of sixteen diverse situations in the culture of Yale students, these sociological findings emerge:

1. No one situation produced overt aggression, strong frustration, or strong conflict in every individual, but every situation did affect someone in these respects. These particular situations, therefore, even though they bristle with cultural regulations, do

[9] It has been shown by Sears and Hovland (5) that the strength of conflicts, as measured by the difficulty of resolution, varies positively with the approach of the strengths of the two conflicting impulses to equality.

not lead to complete uniformity of behavior, although some of them tend to affect a large percentage of individuals similarly.

2. The range of percentages for overt aggression is from 82.4 per cent to 8.5 per cent; for strong frustration from 70.9 per cent to 11.8 per cent; for strong conflict from 53.3 per cent to 21.8 per cent. This range would indicate that strong conflict varies from situation to situation much less than either of the other two factors.

3. Each situation can be described in two ways: it can be characterized in terms of amount of overt aggression, amount of strong frustration, and amount of strong conflict; and it can be compared in one, two, or all three of these respects with the other fifteen situations. Thus it appears that in situation number 8 ("A good friend of yours was driving so recklessly that you were genuinely scared out of your wits") 81.2 per cent of the individuals were overtly aggressive, 70.9 per cent were strongly frustrated, and 30.8 per cent experienced strong conflicts. In comparison with the other situations, this one ranks second in percentage of those who expressed their aggression overtly and first of those who were strongly frustrated, and thirteenth in percentage of those with strong conflicts. In different words, for Yale students this is a very frustrating situation, more frustrating in fact than any other; it makes a large proportion of them overtly aggressive, more of them in fact than all others except one; and it produces strong conflict in only a little more than one quarter of them, which is a comparatively small number in comparison with most of the other situations.[10] In contrast, situation number 13 ("A very casual acquaintance near you cheated on an examination") is the least frustrating of all, it is low in the amount of overt aggression it arouses, and it stands sixth in respect to the strong conflicts it engenders. In a broad, general way it may be adduced that the students have more strong conflicts in this latter situation than in the former, but fewer of them are affected by the sight of someone cheating and they do not (or cannot) express overtly the aggression they feel.

[10] No comparison of ranks or correlations can be made between the three variables, since the denominators vary systematically in the equations giving the percentage values.

CONCLUSIONS

Four hypotheses bearing on factors determining substitute behavior and degree of overt aggression have been deduced from the general laws of frustration and of learning, and evidence gathered from a questionnaire designed to permit standardized recall of real-life situations has been presented to substantiate them. The following generalizations may be made:

1. The frequency of substitute responses varies positively with the strength of anticipatory responses to punishment-for-being-aggressive.

2. The frequency of substitute responses varies inversely with the strength of instigation to the frustrated goal-response.

3. Overtness of aggression varies inversely with the amount of punishment anticipated as a consequence of such behavior.

4. Overtness of aggression varies positively with the strength of instigation to the frustrated goal-response.

Three assumptions have also been tested by these data.

5. Overt aggression is more satisfying than non-overt aggression or substitute responses in the sample frustrating situations used in this study.

6. Direct overt aggression is more satisfying than indirect or displaced overt aggression.

7. There is greater anticipation of punishment for overt aggression than for non-overt aggression or substitute responses in the sample frustrating situations used in this study.

REFERENCES

1. CASON, H. Common annoyances: a psychological study of everyday aversions and irritations. *Psychol. Monogr.*, 1930, **40**, No. 2.

2. DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., AND SEARS, R. R. *Frustration and aggression.* New Haven: Yale University Press, 1939.

3. HOVLAND, C. I., AND SEARS, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23**, 477–493.

4. LEWIN, K. *A dynamic theory of personality.* (Trans. by D. K. Adams and K. E. Zener.) New York: McGraw-Hill, 1935.

5. SEARS, R. R. Effect of relative strengths of conflicting responses on mode of resolution of motor conflicts. *Psychol. Bull.*, 1938, **35**, 664. (To be published *in extenso* as: Sears, R. R., and Hovland, C. I. Experiments on motor conflict. II. Determination of mode of resolution by comparative strengths of conflicting responses.)

6. WICKMAN, E. K. *Children's behavior and teachers' attitudes.* New York: The Commonwealth Fund, 1928.

THE RIGIDITY OF A BASIC ATTITUDINAL FRAME

BY WALTER S. WATSON AND GEORGE W. HARTMANN
Teachers College, Columbia University

The Category of "Framework" Responses in the Systems of Theoretical Psychology

Explanation of many of the "higher" mental processes has insistently demanded the postulation of some controlling pattern such as apperceptive mass, *Einstellung,* determining tendency, delusional system, postural adjustment, sensory orientation, value hierarchy, means-ends readinesses, personality radix, master attitude, etc. Recent discussions in education, politics, philosophy, and the "sociology of knowledge" have independently developed the notions of a frame of reference, pyramid of goods, ideology, or world-model as dominating individual or group interpretation of experiences. Recent field, clinical, and laboratory studies suggest that a fruitful union of these two academic traditions in the inclusive concept of "framework reactions" results in a better understanding of many features of adult behavior, particularly the obscurer phenomena of social psychology.

The chief concrete manifestation of a framework is a guiding pattern of belief and action conveyed by such terms as "theist," "economic radical," "thoroughgoing democrat," etc. Its psychological reality is attested by the evidence of factorial analyses, consistency in individual conduct, homogeneity of opinion within segregated groups, and the internal logic of any system of thought. Persons with an X organization, so to speak, find it difficult or impossible to comprehend those of Y organization. Studies of learning guided by this outlook may indicate that x responses are readily acquired by persons with a preexisting X pattern; y responses, however, will not be easily made

because they fail to fit the frame. Much more resistance must be overcome to alter the *form* of a frame than to modify its *content*. The establishment and disintegration of frames is one of the frontier problems of contemporary psychology. The frame itself is a pure intellectual construct required to unify an extraordinary variety of culturally determined behaviors and is no more to be reified than the other substantives peculiar to psychological discourse.

The Gestalt concept of a mental structure or pattern that acts as a reference field or frame in conditioning local behavior has been placed upon a new experimental foundation since the work of Sherif in his study of autokinetic movement. Sociological investigations and innumerable lay observations had independently noted the remarkable stability and resistance to change of such viewpoints as naturalism, Catholicism, pacifism, and other large-scale orientations as compared to the relative fluidity of specialized learning occurring within the more comprehensive types of "attitudinal frames." Darley speaks of a continuum of increasing resistance to change and notes six points along a rough scale: (1) recently acquired attitudes or opinions to objects remote from the self, (2) opinions based on misinformation or lack of knowledge, (3) superstitions, (4) deep-seated prejudices, (5) opinions or systematic stereotypes based on early home or religious training, and (6) opinions concerning the self.[1] The first three correspond to what we term "intra-field" learning, the last three to "framework" learning. There are, according to our hypothesis, marked qualitative and quantitative differences in the responses thus dichotomized.

The origins of such attitudinal frames and the control of methods for changing them are among the key problems common to education and psychology. We shall, indeed, lack an adequate science of human nature until we are able deliberately to convert a militarist into a pacifist (and *vice versa*) within a reasonable time. Our relative inability to overcome the effects of past learning and to reorganize the drives of the system is probably not a permanent limitation.

[1] Darley, John G. Changes in measured attitudes and adjustments. *J. soc. Psychol.*, 1938, 9, 189–199. The pertinence of this inquiry to the problem of "social lag" is clear.

We may schematize this approach by the accompanying diagram:

A shift from an X-frame to a Y-frame is comparable to the kind of conversion or transformation that makes a "progressive" educator out of a classicist, whereas a person with either structure may readily move from inability to ability to drive a car (represented by a → b). We commonly allege that emotional involvements, self-interest (often confused with self-deception and misinterpretation of what is truly of personal advantage), or early training (negative transfer and inhibition) explain the rigidity of people's attitudinal frames or world-models and their apparent imperviousness to contrary arguments; but this interpretation does not satisfy those cases where a well-chosen and relevant bit of information effectively alters a person's outlook (via the mechanism of insight). We have seen the hostility of some conservatives to the President's Supreme Court proposal melt visibly when they were told that the Constitution does not fix the number of the judges and that their number has been both increased and decreased during earlier administrations.

MEMORY AS AN INDEX OF THE FRAME'S PRESENCE

One of the first questions facing an experimenter studying this problem of altering basic "philosophic" adjustments is the question of the effect of a strong attitudinal frame upon the subject's retention of the material presented to him. If the datum or argument is powerful enough to bend or reorganize a frame, this effect must in some way be made apparent. Would a strong frame cause a subject to reject completely even the most potent stimuli which tended to upset the frame? Or would the subject treat opposing material as Thorndike found "annoyers" were treated? Would there be a significant difference between the recall of relatively innocuous opposing material compared with quite disturbing opposition arguments? Would a series of balanced *pro* and *con* arguments on an issue tend to shift subjects

in extreme positions towards a more neutral position, or would the frame act in such a way that the subject would accept only that portion of the material presented which supported his own original attitude? What modifications would occur in the subject's recall with the passage of time?

The keenly analytical work of Bartlett on remembering implies that *recall* gives a mistaken idea of the process of remembering.[2] A better description may be given by "reconstruction" or "rebuilding," in which the subject starts with one or more cues and makes up out of his own experience a new entity which produces something that structurally resembles the original idea, percept, or feeling. In this process of reconstruction he is necessarily forced to use his present feelings and attitudes as well as previous ones; these appear definitely to warp the concrete material that is remembered. The longer the time interval the greater such warping. Such pattern-distortion is a commonplace of configurationist studies, but its application to problems of belief has seldom been made.

TWO FORMS OF RELIGIOUS ADJUSTMENT

A strong attitudinal frame or existing pattern of belief may be expected to produce significant differences in the overt recall of supporting and disintegrating arguments impinging upon that frame. The area of *Belief in a Personal God* was selected as the attitude to be used in this investigation, because it offered pronounced and well-known attitudinal frames involving a clear-cut yes-and-no polarity between the *pro* and *con* arguments. Groups at the extremes of belief and disbelief were studied. With such groups, any difference in reaction to *pro* and *con* items offered in the same fashion to both believers and non-believers presumably lies in some effects emanating from the subject's attitudinal frame.

Ten theists of varying denominational beliefs were selected from the graduate student body of Union Theological Seminary, the Jewish Theological Seminary, and the Department of Religious Education at Teachers College, Columbia University; to match them ten atheists were chosen from different divisions of the Advanced School of Education in the last-named institu-

[2] Bartlett, F. C. *Remembering*. New York: Macmillan, 1932.

tion. In addition to selection on the basis of reputation and personal statement of belief, the strength of the attitudinal frame was confirmed by the Thurstone Scale No. 22, Forms B and D, and also by a check list of paired-value choices constructed by the authors. The check list contained forty items pairing theist values with other commonly held values, and ten items pairing active anti-theist values against common values. Since this check list was intended to measure the *spread* of the subject's attitudinal frame over a variety of choice situations, it was necessary to include more items for the theist group; because of differing church affiliations, they were expected to have varying concepts of the *values* inherent in their frame. The scope of the atheist outlook seemed to be satisfactorily revealed by the shorter series.[3]

EXPERIMENTAL SEQUENCE OF STIMULI

The test materials were presented to the subjects during an individual interview in a consulting room in the following order:

1. Self-rating on the Thurstone Scale No. 22, B and D.
2. *Free recall of familiar arguments* for and against the existence of belief in a personal God.
3. Rating for *effectiveness* of ten *pro* and ten *con* arguments presented in pairs.
4. Ranking of the same twenty arguments presented singly.
5. Self-rating on the check list.
6. Detailed recall of the arguments presented in steps three and four above.
7. Re-checking of the Thurstone blanks.

Step 1. This step is self-explanatory. It was used in the selection of subjects and as a check on their original position so that any changes from it at the end of the experiment could be noted.[4]

[3] Examples of the check-list items are given here. After such an exhaustive inquiry we can be sure that only *bona fide* theists and atheists were involved.

Theist:

1. You lose your temper without much excuse. Do you feel bad because:
   ..... A friend saw you?
   ..... God saw you?
2. One of two statutes is to be erected in a town square. Would you choose a:
   ..... Religious leader?
   ..... Scientific leader?

Atheist:

1. Your only alternatives are:
   ..... Think as you choose and get $1,000 a year.
   ..... Swear allegiance to belief in God and get $5,000 a year.
2. Your wife requests your permission for a religious burial for your child. Would you:
   ..... Consent?
   ..... Refuse?

[4] Some dissatisfaction with the Thurstone Scales was expressed because of the duplicate or contrary meanings implied in a single statement and also because the concept of deity used failed to match the best modern versions.

Step 2. During the period of free recall the subjects were told that they would be given about fifteen minutes to list as many as possible of the arguments for and against a belief in God that they could recall. They were to make no selection as to importance, but merely to list all the arguments they could think of as briefly as possible. This period of spontaneous recall was included so that the subject's familiarity with the items to be presented to him later might be considered in the interpretation.

Step 3. The twenty arguments for this step were carefully selected from theological and humanist tracts to represent a wide variety of approaches to the problem. The same types of proof or evidence were used in developing opposing conclusions—thus, one pair of arguments based upon authority would be given, another on scientific evidence, another on logical inference from experience, etc.[5]

Each subject was given the same instructions for this step as follows:

> On these cards are ten pairs of arguments for and against the existence of God. Examine each argument of each pair carefully in the order in which they are given to you. After reading a card you will evaluate each argument separately on a three-point scale of effectiveness or importance, either of SUPPORT or ATTACK upon the position you personally hold. Indicate *little* or no importance by "one." Use "two" to indicate the middle position for those arguments which you feel to be of about *average* effectiveness. Use "three" to indicate those arguments which you feel are either very *essential* for your position or which are the opposite, i.e., very disturbing and hard to answer.

It is believed that the chief factor in this situation tending to fix these items in the mind of the subject was his effort to evaluate the material in relation to his own position. The visible purpose of this step was to allocate the material presented into *categories of effectiveness* for and against the subject's position. Recalled materials were later separated into these categories to test differences in memory due to the strength or weakness of the argument.

Step 4. The subject was given the same twenty arguments, each on a separate card, and asked to rank the arguments from one to twenty in a relative series on the same single scale of effectiveness used in the previous step. This process gave addi-

[5] See Appendix I for a complete list of the twenty arguments or ten pairs. Theist and atheist arguments appear equally often in the first and second parts of each pair.

tional time for fixation of the material on the basis of effectiveness and allowed a rough check on individual consistency.

Step 5. The check-list discussed earlier in this paper was included at this point in the procedure to allow for the passage of an interval between the presentation of the arguments and the attempted recall. Although the time elapsing during this step varied from ten to thirty minutes, depending on the person's work rate, it did not affect any comparison between items recalled by the *same individual*. Atheists and theists took both parts of the questionnaire so that the thought stimulation during the "forgetting period" was the same for all individuals.

Step 6. The directions which follow are self-explanatory:

> We appreciate your cooperation up to this point in the experiment, but there is one more important task to be done. This is a surprise step and as such, it is of the utmost importance that you do not tell it to anyone else who might possibly be connected with this experiment. Please recall as exactly as you can and in complete detail each of the twenty arguments you rated and ranked a few minutes ago.

An unlimited time was allowed in which to attempt recall. Subjects varied from ten minutes to over an hour in the time needed to record their memories. All were urged to try to recall fragments even though they could not recall any additional complete units.

Step 7. The subjects were given their *original* Thurstone sheets to re-check with the following instructions:

> In the light of all the stimuli, experiences, and thoughts that have occurred in the last hour, please re-check these scales indicating your new position by appropriate marks at the left of the parenthesis.

No attempt was made to conceal the former marking on the scale. Frequently subjects asked if they should indicate alternative markings that were in mind when the scale was originally checked. They were told to indicate only changes which they felt had taken place since the first marking of the items. Although almost no items were changed, it was obvious that the presence of earlier markings had a tendency in most of the subjects to give undue weight to the "no change" response. It did, however, minimize the purely accidental variations which the presence of alternative responses in the first marking might have produced.

ANALYSIS OF THE DATA

Each item recalled under Step 6 above was carefully studied in comparison with the material originally presented to the subject. After identification with the original stimuli the recalled items were recorded under six categories of effectiveness corresponding to the individual's rating assigned to the argument in Step 3. Thus the first item recalled by subject X corresponds with argument 15; in step three he had also assigned a value of 2-*pro* to this argument. This recalled item was listed as a recall of *pro* material of average effectiveness. All other remembered data were similarly classified.

Two further distinctions were made among the recalled items. First, the items were given a *quality* score by the examiner ranging from one to four. The quality comparison was kept *within* the individual's performance range so that a 3 or a 4 quality score meant the best work of an individual and a 1 score his poorest; in a few cases the work was so uniformly good or so consistently bad that the entire range did not appear. Second, a word count was made and compared with the number of words actually used in the argument the subject was attempting to recall. Such a count merely furnished a measure of comparative verbosity on the different items.

For each individual three indices were thus secured: (1) the *number of items he recalled* under each of the three distinct *pro* and three *con* categories of "rated effectiveness"; (2) the *quality* of his responses; and (3) his *verbosity* during the reproduction.

COMPUTING THE MEAN RECALL VALUE FOR EACH CATEGORY

In obtaining the mean memory value of each category special consideration had to be given to items which the individual had listed in his complete rating of all arguments under step 3, but which he failed to recall. Two opposing hypotheses are possible:

(1) It may be argued that since no one can know what fraction of the items omitted in the overt recall was subliminally present ready to be released upon "proper" stimulation, the data should be treated solely on the basis of such items as were recorded.

(2) On the other hand, no recall is certainly worse than a fragmentary word or two which were assigned to the "one" category of quality; prompting or other help applied to such an argument might have enabled the subject to boost his rating to a "three" or "four" quality. Consequently, items not remembered in the overt response should be given a rating less than one—that is, zero.

Most of the means were obtained on the basis of the second consideration; i.e., items that were not presented by the subject in his attempted recall are listed as zero responses. Where the means are based on the *items positively recalled* exclusively, they will be so labeled.

Procedure for Determining Significant Differences. Snedecor's "F" based on Fisher's "Z" was used to test for significance the differences between groups of means. This procedure avoids the labor of reducing the material to standard scores for the purpose of eliminating the variation among individuals and also makes proper allowance for the distributions to be expected where the number of cases is small.[6]

The null hypothesis was rejected and findings labeled as "significant" only where the differences could have been produced by chance in less than one case in twenty out of a series of random samples of the same population. Where the chances of an accidental difference as large or larger than the observed difference are less than one in a hundred, the results were called "very significant."

The Issue of Familiarity. Any unequal distribution of familiarity could well have produced spurious differences in recall. A check on the material given in Step 2 (the period of initial free recall) shows that 80 *pro* arguments and 75 *con* arguments were offered by the two groups taken together. The atheists listed 48 *pro* and 38 *con;* the theists 32 *pro* and 37 *con.*

Upon comparing these familiar items with those presented to the subjects for later recall, however, we find that only 32 *pro* and 32 *con* referred in any detail other than conclusion to the content of the arguments. The A's were now divided 25 *pro* and 15 *con* and the T's 7 *pro* and 17 *con.* These numbers

[6] For a discussion of this useful technique and for tables of F upon which our conclusions were based, see Snedecor, G. W., Calculation and interpretation of the analysis of variance and covariance, Ames, Iowa: Collegiate Press, 1934.

refer, of course, to the total frequency with which familiar arguments were mentioned by all the subjects and not to the actual varieties of *pro* and *con* arguments.

While it is true that we have no measure of the possible effect of arguments which were not recalled in Step 2 but which were later recognized by the subject as familiar, the small proportion mentioned of the 400 possible (i.e., 20×20) led us to reject the hypothesis that our observed differences were primarily due to differences in familiarity, particularly since our method rests essentially upon intra-group comparisons.

MAIN TABULAR REPORTS AND DISCUSSION

Table 1 presents the average *quality* of recall (as judged by the experimenter) of items in the three categories of effectiveness assigned to the arguments by the subjects. For complete details of this method of calculation Appendix II should be consulted.

TABLE 1

AVERAGE QUALITY OF RECALL (INCLUDING ZERO RESPONSES) OF ARGUMENTS CONCERNING GOD'S EXISTENCE

(Mean Recall Quality for Three Categories of Rated Effectiveness)

GROUP	1	2	3	
All Subjects	1.28	1.38	1.54	Differences Significant (S)
Atheists	1.66	1.53	1.78	Differences Not Significant (N.S.)
Theists	0.90	1.24	1.31	Differences Significant (S)

Table 1 should be interpreted as follows: 1.28 is the average quality of recall for all subjects as rated by the experimenter on those arguments which they had placed in the "one" ("unimportant") category of effectiveness. Arguments originally deemed of moderate effectiveness by the subjects were given a mean rating of 1.38 for quality of recall when reproduced by the subject and judged by the examiner on a 4-step scale. Those in the "three" ("highly important") category of rated effectiveness for arguments either supporting or attacking the subject's position were recalled with a quality of 1.54. The differences among these means are considered significant (S), which means

that they are greater than those occurring by chance in 95 per cent of random samples from the same population when groups of this size are used. A similar interpretation is to be given to the two succeeding rows, which show how the two groups of subjects behaved when considered independently. The figures suggest that the intrinsic strength or power of an argument can be recognized by both friend and foe.

TABLE 2

AVERAGE QUALITY OF RECALL OF REASONS FAVORABLE OR UNFAVORABLE TO ONE'S BELIEF

(Mean Recall Quality for Three Categories of Rated Effectiveness)

GROUP	PRO			CON			
	1	2	3	1	2	3	
All	1.42	1.42	1.59	1.16	.94	1.37	(N.S.)
	2 AND 3			2 AND 3			
	1.47			1.17			(N.S.)
	1	2	3	1	2	3	
Atheist	1.94	1.69	1.76	1.41	0.73	1.87	(S.)
Theist	0.69	1.15	1.42	0.92	1.08	1.09	(N.S.)
	2 AND 3			2 AND 3			
Atheist	1.66			1.22			(N.S.)
Theist	1.28			1.09			(N.S.)

In Table 2 we present the means for various categories of rated effectiveness with a finer division than above, separating the arguments favoring the subjects' attitudinal frame *(pro)* from those which opposed it *(con)*. As before, the scores represent quality of recall as rated by the experimenter; omitted items are included as zero in determining the means.

Although but one of the sets of means is statistically reliable, it will be noted that in all cases the "three" ("important") items were better recalled than the "two" items and that the combined "two" and "three" *pro* items were recalled better than similar *con* items.

Table 3 is similar to Table 2 except that the means are now based solely on such material as was *positively* recalled by the subject. Neither set of means differs significantly.

TABLE 3

AVERAGE QUALITY OF RECALL WHERE NON-RECALLED ITEMS ARE ABSENT FROM THE CALCULATION

(Mean Recall Quality for Three Categories of Rated Effectiveness)

GROUP	PRO			CON			
	1	2	3	1	2	3	
Atheist	2.61	2.57	2.62	2.31	1.87	2.12	(N.S.)
Theist	2.06	1.99	1.98	1.78	2.24	1.62	(N.S.)

In Table 4 we attempted to check our hypothesis in terms of the *gross number* of arguments recalled by the subject in each category of effectiveness, regardless of the quality of his memory. The scores represent the percentage of items which the subject recalled. Thus, if an atheist had placed four of the ten atheist arguments in category "one" of effectiveness and recalled portions of only two of these arguments in his attempted recall, he received a score of 50 per cent for his items in the "one" category.

TABLE 4

MEAN PERCENTAGE OF ARGUMENTS RECALLED IN EACH CATEGORY OF EFFECTIVENESS

(Average Per Cent of Recall for Three Categories of Rated Effectiveness)

GROUP	PRO			CON			
	1	2	3	1	2	3	
Atheist	65	63	65	55	34	87	(V.S.) [7]
Theist	32	52	71	71	55	60	(N.S.)
	2 AND 3			2 AND 3			
Atheist	62			54			(N.S.)
Theist	63			60			(N.S.)

[7] Although the percentages given for differences within the means of the atheist group are labeled statistically significant it appears on inspection that the figures for the "two" and "three" Con categories are responsible for the bulk of the variation. A glance at Table 5 which follows will show that so few items were placed in these two categories by the atheist group that we do not have a reliable index in these two columns, and hence must reject the findings as significantly different from chance for the table as a whole.

Table 5 was constructed to show the results of Step 3 in the procedure, so that the persuasiveness of the material used might be noted. It was thought that the experimenter may have favored either Theists or Atheists in the number of effective arguments selected. In the upper half of the table the means are arranged under the three categories of "effectiveness" and the two classes of *pro* and *con*. In the lower half of the table, atheist arguments and those for theism are separately presented with the average number of arguments placed in the three categories of effectiveness.

TABLE 5

AVERAGE NUMBER OF ARGUMENTS PLACED IN EACH CATEGORY OF RATED EFFECTIVENESS BY THE SUBJECTS (SUMS OF ROWS = 20)

GROUP	PRO			CON		
	1	2	3	1	2	3
All	2.25	3.40	4.35	6.45	1.75	1.80
Atheist	2.30	3.40	4.30	7.40	1.50	1.10
Theist	2.20	3.40	4.40	5.50	2.00	2.50
	ATHEIST ARGUMENTS			THEIST ARGUMENTS		
	1	2	3	1	2	3
All	3.60	2.70	3.40	4.80	2.45	2.75
Atheist	2.30	3.40	4.30	7.40	1.50	1.10
Theist	5.50	2.00	2.50	2.20	3.40	4.40

Table 5 is read as follows. When the atheists ranked the ten arguments supporting their position, they placed a mean of 2.30 arguments in the unimportant or "one" category, 3.40 in the class of "average" or moderate effectiveness, and 4.30 in the "very important" category. This same group placed 7.40 arguments on the average in the *unimportant* category when considering opposition material, 1.50 in the moderately effective group, and 1.10 in the highly effective group. The theists when considering the arguments that supported their views showed the same pattern of distribution, but were more lenient with the "opposition," as revealed by the 2.00 and 2.50 in the "two" and "three" categories.

In the lower half of the table we have a rearrangement of the same figures appearing in the upper half except for the row labeled *all*, which combines the two groups. Here we can see clearly that the atheists thought that 4.3 of the atheistic arguments were very important, while the theists felt that only 2.5 were very important. When judging the theistic data, the atheists felt that only 1.1 arguments in the ten were very effective in attacking their atheistic position, while the theists held that almost half, or 4.4, were very strong arguments in favor of their theism.

We have already discussed the distribution of quality ratings by the examiner in Tables 1, 2 and 3, and in Table 4 we interpreted the meaning of the *number* of arguments recalled. A

TABLE 6

MEAN RATIO (IN PERCENTAGE) OF WORDS USED IN RECALL TO THE WORDS USED IN THE ORIGINAL ARGUMENT (NOT INCLUDING ZERO RECALL)

GROUP	PRO			CON			
	1	2	3	1	2	3	
Atheist	39	43	47	41	34	33	(N.S.)
Theist	17	32	30	35	26	26	(N.S.)
	2 AND 3			2 AND 3			
Atheist	47			26			(V.S.)
Theist	30			28			(N.S.)

third distribution of the scores was made upon the basis of verbosity in the categories of rated effectiveness. *Verbosity* was roughly determined by the ratio between the number of words used when recalling an item and the number appearing in the argument as originally given to the subject. As in the previous indices, we are concerned here not with the differences between individuals, but with differences in treatment among the categories by the same individual.

The only significant difference in the use of words is the fact that the atheists' *pro* of 47 per cent exceeds the *con* of 26 per cent by a margin greater than one would expect by chance.

Chance Factors in the Results. Statistically all variations in recall which were not identified with our hypothesis have been

dealt with as *chance factors.* Theoretically this procedure is not justified. We know in some cases that factors were present which tended to reduce the observed differences, particularly the presentation of arguments in related pairs for procedural simplicity. This pairing tended to associate *pro* with *con* arguments (via closure and other learning mechanisms); if one were recalled, it brought the other in its wake. Pairing also reduced the differences between strong and weak arguments, because of the tendency of subjects to say that if argument X, resting on Y type of proof is *strong,* then the associated argument Z—also resting on Y type of proof but reaching the *opposite* conclusion—must be *weak.* Since, however, this pair of strong and weak arguments was tied together by the common bonds of joint presentation and apparently identical type of proof, if one were remembered we might expect its associated partner to arise. In fact, some subjects spoke orally during their recall period of using this association to dig up material which they found difficult to remember.

Other elements treated as chance were partial familiarity with the *content* of the illustrations, terminology, etc., and items which stood out prominently, such as the one argument containing "statistics." Other arguments were retained (according to subjective reports) because they represented a neat, integrated format, as is the case with slogans and jingles. Others had emotional connotations with situations or individuals in the subject's previous experience. An example is a subject who recalled an "unimportant" argument because he had been so disgusted when he heard that type of reasoning used to attack his position. We have no way of estimating the extent or direction of most of these incidental factors in which the content was clothed except those produced by alternative pairings, which it is apparent tended to *reduce* the differences between categories.

Re-shaping of Content in Recalled Material. There were occasional bits of evidence in the recalled material of a tendency to re-shape the opposing arguments in such a way as to make them less damaging to the attitudinal frame held by the subject. For example, the following material was presented as Argument No. 15:

A careful investigation by Leuba shows that scientists who are ranked by their fellows as outstanding in their field have less tendency to believe in God

than do rank-and-file scientists in the same field.[8] Also, those scientists who know most about men and their values have less belief than natural scientists.

PERCENTAGE OF BELIEVERS IN ANY FORM OF PERSONAL GOD

	PHYSICISTS	BIOLOGISTS	SOCIOLOGISTS	PSYCHOLOGISTS	ALL SCIENTISTS
Lesser	43	31	30	13	35
Greater	17	12	20	2	13
Greatest	..	..	5	..	..

Compare with this the three sets of recalled material given by three theist subjects:

(1) "As far as belief in a personal God—It has been shown that physicists show a greater degree of credulity than other scientists and 5 per cent of the greatest sociologists do believe in a personal God."

(2) "Somebody investigated the belief of scientists in God and found that few of the most eminent believed in God, but many of the 'general run' did. Five eminent sociologists had a very firm belief in God, the only group of eminent scientists so listed."

(3) "A lower percentage of scientists considered by L . . . . . to be greater showed a disbelief in God as compared to the mediocre scientists."

Assuming that the complete re-centering of the third answer may have been a slip of language (shades of Freud!), one may see in these answers a type of response similar to the story of the mate who, hating his captain and having a chance to write up the ship's log, recorded quite truthfully, "The Captain was sober today."

Note the changes in the following recall of an atheist subject. He has introduced a humorous element, omitted the chief point of the argument—the implication of a Deity—and added a "Sic!" at the close as if he were saying, "I was only quoting, I don't believe this."

Original Argument:

The very fact that there is so much pain, evil, and suffering in the world makes it obvious that life wouldn't be worth living if there were not a God who can eventually recompense those who have suffered unjustly in this life.

As Recalled:

This is such a tough old world surely there must be some compensation in after-life (Sic!).

[8] Leuba, James H. Religious beliefs of American scientists. *Harpers*, 1934, **169**, 291–300. This source reference was not given to the subjects.

CONCLUDING REMARKS

The effect upon recall of pertinent arguments of strong theistic and atheistic attitudinal frames was studied. For the group under investigation, material subjectively judged to be effective was recalled in greater *extent* and with better *quality* than that rated as having average effectiveness, and material which *supported* the subject's attitudinal frame was *retained better* than material which opposed it.

Although the results as a whole consistently buttress these conclusions, most of the differences taken singly *are not statistically significant.* In the large picture, however, the general direction of the differences reinforces the hypothesis. The limited number of subjects and presentation of the material in pairs of arguments are probably major factors concealing bigger differences. A further investigation of the effects of longer periods of forgetting would also seem advisable in view of the results of Bartlett's work where the interval between stimulation and recall and the amount of change or modification appear to be positively correlated.

This account of a prolonged inquiry seems excessively wooden and uninspired. The experimenters end it with a marked sense of frustration. We had hoped to be able to demonstrate some effect of information *per se* (or of knowledge of the opponents' reasoning) in bending such a rigid set as one's religious outlook. One cause for man's faith in what is termed logic and scientific method lies in the belief that when all the essential facts about a controversy are assembled, these facts themselves will spontaneously point to one conclusion rather than others. In the face of such evidence even the most powerful predisposition presumably must yield. It seems, however, that either most great issues are like ambiguous perceptual figures which permit more than one pattern of apprehension to emerge, or the grand frames which encompass our mental life are reconstructed only when the organism experiences some powerful trauma. With all the methodological imperfections of this study, these findings give scant comfort to the liberal and the educator. At the end of a two-hour period of earnest response to alternatives there had been no perceptible approximation of the frames of the opposed groups. Does this fact mean that we are permanently doomed

to "temperamental incompatibility" and "ideological pluralism"? Or are the dynamics of fundamental transformation so subtle that they elude us?

APPENDIX I

The arguments used for evaluation and recall in Steps 3, 4, and 6 of the experiment are listed below. In Step 3 the arguments were presented in pairs (designated by the Roman numerals). In Step 4 they were presented separately on individual cards (indicated by the arabic numerals) but in the *same order* as in Step 3.

I. (1) The creative work of the world has been done by those who believed in God and found their support against a hostile world in their feeling of having the backing of the Order of the Universe; of a God whose instrument they were; of a Something greater than Self—a Something of permanent value which would conserve their effort toward the Good. Even Communists today, have such an abstract ideal—"Historical Destiny." The history of all human bases for social effort is one of disillusionment. Only in an "at-one-ment" with God can the flood gates of man's creativity be loosed in such a way that they will not be dried up by an ultimate cynicism when things go wrong.

(2) Many of the principles of action proposed by the great men of the past were truly sublime. But the concept of God which they followed led to an "Other Worldliness" which has been the opiate of their followers ever since. It has caused them to shut their eyes to the evils and injustice of the present in a vain hope of happiness and justice in some future life with God. People do need a Cause to claim their devotion, but this cause may be entirely earthly in origin. In fact, the closer the connection between the Cause men live by and the realities of earth, the greater the chance for satisfactory fulfilment of their desires.

II. (3) Theism argues under various guises that behind every chain of events there must be a starting point, a first cause. Atheism answers, *Why?* Who or What caused God? Why shouldn't there be an infinite series of antecedents which never end?

(4) Everything including the Universe has a cause; and though men differ about the nature of the cause of the Universe (this Spring from which flow the Rivers of Life) none can deny that there is such a cause and most would call it God.

III. (5) Deep in man's heart lies a hungering, insistent yearning for God. When man acts as if God exists, this need is filled. The only possible conclusion one can draw from this fact is: God exists.

(6) Man's explanation of why certain action brings relief to his felt needs for a God may be as false as the heathen explanation of why his prayers addressed to a stone image are followed by rain.

IV. (7) There is an observed order, causality, and predictability in the occurrence of natural events; but why must there be a divine mind or God? The relation of parts to the whole is purely a naturalistic fact that needs no personal Deity to account for it. The analogy usually drawn from the ability of the mind to create new patterns of order is no more valid than to argue that since light is the source of energy and life, God must be like some infinitely big star.

(8) If so small a piece of mechanism as a watch could not come into existence without an intelligent designer and maker, all the more would this be impossible for so elaborate and orderly a structure as the whole universe. This universe antedates and embraces us—we are mere parts in the larger total. Man has not originated the universe nor its laws. The creator of this larger whole is God.

V. (9) The scientist in his laboratory, seeking to conquer some insidious disease; the social worker laboring to correct the maladjustments of family life; the missionary giving his life for an alien people—all progressive activities testify to the presence of values in the sum total of Reality. But Inanimate Creation shows no interest in such struggles, and mere mechanism is indifferent. All sublime aspirations speak to us of a God in whose life they are fully realized and from whom these come into the heart of man.

(10) Acceptance of God should lead to progress and understanding if he really exists. History shows exactly the opposite to be the case. The real progress of the human race away from an early animal existence has depended upon those great thinkers who were able to deny the God of tradition and do their own thinking. Control over growth of plants, rivers, electricity, animal breeding—in fact, anywhere that progress has been made—has meant denial of traditionally accepted divine control.

VI. (11) Two thousand years ago Epicurus said, "Either God wishes to destroy Evil and cannot; or he can but will not; or he neither wishes nor has the power; or he both desires and is able. If he wishes and cannot, he is impotent; if he can but will not, he is wicked; if he can and will, *why does evil still exist?*"

(12) The very fact that there is so much pain, evil and suffering in the world makes it obvious that life wouldn't be worth living if there were not a God who can eventually recompense those who have suffered unjustly in this life.

VII. (13) One of the few points that all races both primitive and modern hold in common, however crude or refined they may be, is some idea of God. Surely if the concept arose only through an error in man's thinking, we wouldn't find it so universally among races separated through ages of history.

(14) Theists talk of the universality of the idea of God. Actually there is perhaps no term in common use which has so many different meanings. God is a term which represents the unknown in man's mind. No two men and no two religious sects have the same unknown or the same God. One religion sees Him concerned with storms, another with vengeance, another with love, another with man's lusts, another with chastity, etc. A term so all-embracing borders on the meaningless.

VIII. (15) A careful investigation by Leuba shows that scientists who are ranked by their fellows as outstanding in their field have less tendency to believe in God than do rank-and-file scientists in the same field. Also, those scientists who know most about men and their values have less belief than natural scientists.

(16) A short time ago, "scientists" were telling us that the world was a mechanism—no purpose—no God—man was but a machine. Today they are presenting the sanctity of their latest discovery—the world is an organism—there is no such thing as matter—goal-seeking and purpose are the mark of a human. Shall we change our religion every time they discover they were

wrong before? Or shall we recognize that they are still lost in the dogmatism of their own "scientific method" and exclusive "arm-chair" philosophizing; that they are on their way to an agreement with truths greater men discovered intuitively ages ago and the common man discovers in the swelter of everyday living? Why worry about what the scientist says about God since he is still experimenting? In five more years he himself will have discovered a new answer.

IX. (17) When a man reads a book or studies a musical composition, there is a strong presumption that there is something in common between the intelligence that created the book or the music and the intelligence that understands it. So as man reads the Book of Nature, he infers that behind it there must be a Creator with an Intelligence akin in some way to his own.

(18) The above argument is just as senseless as to say that since eyes are of supreme importance to read and understand books, then books must in some way be like eyes.

PERCENTAGE OF BELIEVERS IN ANY FORM OF A PERSONAL GOD

	PHYSICISTS	BIOLOGISTS	SOCIOLOGISTS	PSYCHOLOGISTS	ALL SCIENTISTS
Lesser men	43	31	30	13	35
Greater men	17	12	20	2	13
Greatest men	..	..	5	..	..

X. (19) The argument for God resting on men's direct experience of the feeling of another presence is false. Such an experience may fairly easily be produced by known sets of conditions not dependent upon the existence of any spiritual being. Hypnotic states, emotional needs, suggestion, drugs, etc., all produce in individuals at times a feeling of such direct experience of another being.

(20) If man will truly seek God, he will experience the sense of another being in just as real a fashion as if he sees his neighbor with his own eyes. Tides of power beyond our margins sweep into us and refresh us when we contact this being. Moses received the ten commandments from direct contact with God. Christ talked with him often in prayer. Even the common sinner, if he truly repents, knows the reality of contact with the Divine.

APPENDIX II

Space does not permit the inclusion in this article of all the raw data nor a description of most of the intermediate steps involved in reaching the final results as tabulated. For clarity's sake, however, the procedure used in obtaining the first *row* of means in Table 1 is included below.

The primary purpose of this analysis was to determine the variations in quality of recall as the subjects tried to remember arguments which they had listed as unimportant compared to those which they had rated as of medium effectiveness or very effective.

For this reason the Quality Scores assigned by the experimenter to the subject's efforts to recall various arguments were distributed among the three classes of effectiveness as previously rated by the subject as follows:

SAMPLE I (FOR SUBJECT #1)

QUALITY OF RECALL SCORES DISTRIBUTED UNDER THREE CATEGORIES OF RATED EFFECTIVENESS (R.E.)

ORDER OF RECALL	QUALITY OF RECALL		
	R.E. 1	R.E. 2	R.E. 3
1			3
2			2
3			3
4	2		
5	3		
6	3		
7		3	
8	1		
9		2	
10	3		
11	2		
Sum	14	5	8
Number recalled	6	2	3
Number omitted	3	2	4
Total Number	9	4	7
Mean (without omissions)	2.34	2.50	2.67
Mean (including omissions)	1.56	1.25	1.14

Then for each of the other 19 subjects similar mean values were determined and recorded in Table 1. Once the mean score for an individual was secured, it was used as a unit of *equal* weight in combining it with the mean scores of other individuals, regardless of the *number of items* employed in calculating each mean. This technique seemed necessary to avoid weighting the group means on the basis of a few individuals who recalled several times as many items as the poorer members of the group. It does, however, produce apparent arithmetical discrepancies when columns are combined.

The "variance" of these three column means *about* their mean when compared to the variance of the scores *within* the columns (i.e., around their respective "columnar" means) gave an F of 3.69; this may be compared to an F expected to occur by chance in 1 per cent of samples of such size of 5.18 and to an F .05 of 3.23. Since the *observed* F is larger than F .05, it is called "significant." If it had been greater than F .01 we should have called it "very significant."

These two samples itemize the procedure for obtaining the first row of means in Table 1. The other figures in Table 1 (as well as in the remaining tables) were obtained by the same method.

SAMPLE II

AVERAGE QUALITY SCORE (INCLUDING ZERO RESPONSES) GIVEN BY THE EXAMINER TO THE ARGUMENTS RECALLED BY THE TWENTY SUBJECTS LISTED UNDER THE THREE CATEGORIES OF RATED EFFECTIVENESS (R.E.)

SUBJECT NUMBER	AVERAGE QUALITY SCORE			
	R.E. 1	R.E. 2	R.E. 3	
1	1.56 [9]	1.25 [9]	1.14 [9]	
2	1.25	1.20	2.00	
3	2.12	0.25	1.43	
4	2.80	2.80	3.20	
5	1.67	1.67	1.50	
6	1.00	1.00	1.10	
7	1.64	3.00	1.88	
8	0.00	0.50	0.33	
9	2.00	2.60	2.25	
10	2.57	1.00	3.00	
11	0.25	1.00	1.00	
12	2.43	2.20	1.62	
13	0.67	0.72	2.00	
14	0.60	1.71	1.33	
15	0.72	0.20	0.87	
16	1.11	1.50	1.22	
17	0.78	1.17	1.60	
18	1.46	1.67	2.25	
19	1.00	1.25	1.00	
20	0.00	1.00	0.17	
Sum	25.63	27.69	30.89	
Average	1.281	1.384	1.544	[Differences Significant (S)]

[9] The raw data from which these three means for Subject 1 were obtained are presented in Sample I above.

THE FUNCTIONAL SIGNIFICANCE OF CHANGED RESPONSES IN A PSYCHONEUROTIC INVENTORY

BY MAX HERTZMAN
College of the City of New York
AND ROSALIND GOULD
Brooklyn College

The Problem

When a personality inventory is taken on two different occasions by the same group of subjects, the items are not answered in exactly the same way on each testing. An item which is marked differently on the two testings may be called a changed response, and the total number of such responses for each subject may be called his "change score." The main problem in the present study was to find the relationship existing between the change scores of the items of a typical psychoneurotic inventory and some other important features of those items. The other features were the frequency with which an item was answered in the neurotic direction by the test group as a whole and the discriminating capacity of the items. A secondary problem was to study the influence of the form and character of the items on the change scores.

History

A number of studies have appeared in the past few years concerning the changes in responses to items in psychoneurotic inventories. Some of these studies, such as those of Benton and Stone (3) and Neprash (14), have dealt with the gross number of changes in response as a function of the time between testing. These investigators have found that a constant percentage of change is reached very early, Neprash finding no significant difference between a two- and an eight-week period (his extremes), Benton and Stone finding no changes in average consistency scores between 4 and 21 days (the latter being the longest interval studied). Over long-time periods there also may be some sort of stabilization. Farnsworth found that on the

Bernreuter "71.35 per cent of the items were answered identically after one year, 65.45 per cent after two years and 64.91 per cent after somewhat over three years" (5, 11). The gross amount of change found in the above studies when the time intervals were equivalent was not identical, however, and it is likely that the total percentage changed is a function of the inventory as well as of the subjects, with age, sex, education, and socio-economic status important variables which must be considered in connection with the subjects.

Most studies of change have been concerned not so much with the gross total of changes as with a comparison of changes of different kinds of items. Bain (1), in a study of college freshman, found an average of 27 per cent change on items relating to factual family data, 24 per cent change on subjective personal data and 18 per cent change on factual personal data. It is interesting to note that some of the responses to factual items had been changed more frequently (although not reliably so) than some of the subjective items. Smith (16), repeating Bain's work, found a higher change than did Bain on subjective personal items. Frank (6) studied the change of responses on the Bernreuter in a group of males, averaging 23 years, who were somewhat older than the subjects utilized in the previously mentioned studies. Unfortunately, the group to whom the retest was given consisted of only 35 individuals. The items were divided into positive, negative, and neutral items, depending on whether the predominant answer was "yes" or "no." Those items for which fewer than 75 per cent of the responses were either "yes" or "no" were considered to be neutral. Although the index of change was similar for the "yes" and "no" items, it was about 10 per cent higher for the neutral items, averaging 32.5 per cent per item.

Changes of responses have often been regarded as indicating the unreliability of personality inventories, this unreliability being somewhat mitigated, however, by the fact that there may be compensatory changes which decrease the influence of changed responses on the total score. In accordance with this attitude, some have recognized the existence of changed responses without attributing much importance to them. Frank criticized Symonds (17) for adhering to the latter viewpoint. "Clinical

experience . . . does not support such a position. In fact the situation is quite the reverse and only a qualitative analysis of specific item response, will give the clinician a point of approach to each problem" (6, 324).

Certainly it would be important to determine why there are changes and to what structural and functional features of the items they are related. We have already seen that the matter of whether the items are subjective or factual might influence the results. Changes found by Hollingworth (9) have been attributed by him to motivational factors. He reports a different ratio of unfavorable responses by neurotic soldiers to the Woodworth Personal Data Sheet before and after the armistice, with a marked dropping off of neurotic responses following the armistice. Landis (11) has suggested that part of the unreliability of the tests may lie in the fact that some questions are open to more than one interpretation. If an individual answers at different times in accordance with different interpretations, it is possible that a change in the response will be brought about. Landis and Katz (12), comparing the answers given by groups of psychotics and psychoneurotics with their case histories, found many discrepancies. The nature of these discrepancies may provide a cue, however, to the reason for changed responses to items on the inventory. Among the relevant reasons found for the difference between case histories and answers to the questions are carelessness, differences in interpretation of frequencies implied by the question (e.g., "Do you often feel just miserable?"), recovery of the patient, and differences in definition.

Studies of the relationship of change to various functional features of the items are few. Benton (2) devised two measures, which he calls the index of scatter and the ambiguity of an item, respectively. The measures of scatter and ambiguity are related to some of the above-mentioned factors. The index of scatter was obtained by subtracting the frequency of the modal interpretation of an item from 100. This method would mean that the higher the index of scatter, the greater the number of individuals who disagreed with the most frequent interpretation of the item. The measure of ambiguity was the frequency with which a group of 40 subjects spontaneously noted more than one interpretation of an item. These measures were correlated with

the critical ratios of each item, obtained by evaluating the difference for each item between a group of normal and abnormal subjects. The correlations obtained were insignificant, being .10 with the index of scatter, and .01 with the measure of ambiguity.

Neprash (**14**) presents correlations ranging from .40 to .60 between the neurotic scores of his subjects and the number of their changes. He interprets this finding to mean that the responses of the less well-adjusted subjects tend to be less reliable than those of the better adjusted. This interpretation is somewhat in conflict with the warning issued by Landis and Katz (**12**), which points out the possible lack of validity of the low scores obtained on these inventories. Neprash considers a change in response to be an index of the lack of reliability of the questionnaire. He mentions, with hardly any comment, that 24 of the Thurstones' (**18**) most differentiating items, 21 out of 30 of Harvey's (**8**), and 16 out of 25 in a list selected by Willoughby (**19**) are in the upper quartile of unreliability. That an item which is more differentiating should be also more unreliable seems a contradiction. What has actually been discovered is that those items which have been changed most frequently tend to be the better discriminators. Items which are not frequently answered in the neurotic direction do not call for changes on the part of the subject. The responses for items which in the general population may be highly diagnostic or appear frequently as symptoms, are more likely, therefore, to be altered.

PROCEDURE

1. *Psychoneurotic Inventory.* The inventory used consisted of the 42 items selected by Garrett and Schneck (**7**) from the original Woodworth Personal Data sheet (**20**) on the basis of their discriminatory power. Twenty-five of the items showed 10 per cent or more than 10 per cent unfavorable answers by the psychoneurotic group of 256 subjects as compared with the group of 103 male freshman with whom they were contrasted. The remaining 17 were answered unfavorably by at least 20 per cent of the psychoneurotics.

2. *The Subjects.* The subjects used in this study were 147 women students, 97 per cent of whom were in their freshman year at the New Jersey College for Women. The average age

at the first testing was 18 years and 6 months, with a standard deviation of 13 months.

3. *Method of Testing.* A sheet containing the 42 items was presented to each subject. They were instructed to answer "Yes" or "No" to each of the items. The inventory was given again 4 weeks after the original testing. The problem of the optimum interval between testing involves the consideration of many factors. Whereas a short interval may result in the subjects utilizing their memory of their previous responses, testing after a long interval may be influenced by far-reaching changes in the subjects. Our retest occurred 28 days after the first testing, a time which falls well into the period found by Neprash (14) to yield a constant number of changes.

4. *Scores.* For both the individual subjects and the individual items the first 5 of the following scores were obtained. (a) and (b) First, there was the number of unfavorable responses on the first and second test, which will be called A and B scores, respectively. (c) There was also the number of responses which were in the unfavorable direction on at least one of the tests, which will be called T scores. If an item was answered in the unfavorable direction on both testings, or on the first testing alone, or on the second testing alone, it contributed to the T score. Items which were answered unfavorably in both tests A and B were not given additional weight. (d) The number of changes of response for each subject and the number of changes for each item were called the C scores (change scores). The total scores were not obtained directly but were derived from the relationship existing among the scores for the first testing (A), the second testing (B), and the change scores. In order that T be found from the above-mentioned quantities, we must be able to find the number of items which were answered in an unfavorable direction in both A and B. If we let the number of items marked in the unfavorable direction both times equal S (same scores), then

$$(1)\ T = S + C$$

$$(2)\ C = A - S + B - S = A + B - 2S$$

$$(3)\ S = \frac{A + B - C}{2}$$

$$(4)\ T = \frac{A + B + C}{2}$$

(e) Same scores (S) as defined above in (d) were obtained. (f) The difference between the number of unfavorable responses on the two testings was obtained for each subject and was called D (difference score). (g) In order to study the relation of some of the above-mentioned scores to the discriminatory values of each item, critical ratios were obtained for the A, B, and T scores. The 25 highest and 25 lowest individuals were selected and the ratio of the difference between the contrasted groups for each item and the standard error of the difference was obtained, making 3 such ratios in all for each item. Whether the selection was on the basis of A, B, or T, the composition of the groups was not markedly different. The calculation of the critical ratios was facilitated by the use of the Edgerton and Paterson tables (4) for the standard error of a percentage.

RESULTS

1. *Reliabilities.*

(a) Split-half and test-retest coefficients. Although A and B were the same test, taken at different times, their split-half reliability coefficients are not the same. Examination of the coefficients in the light of their predictabilities above chance corroborates this difference. The reactions of the subjects as a whole to the second testing differed somewhat from their reaction to the first testing. There is some difference in central tendency, although it is not reliable. The mean unfavorable response for A was 9.95; for B it was 9.05. The critical ratio obtained by dividing the difference by the standard error of the difference is 1.77. The measures of variability, the standard deviations, are 4.71 and 4.97, respectively. The results of the second testing seem to be more reliable. In connection with another questionnaire, Lentz (13) has also found that the split-half reliability of the second testing was more reliable than that of the first. The reliability coefficient of B and the test-retest coefficient are practically identical, being .875 and .878, respectively.

The split-half coefficient obtained for T is relatively low, the uncorrected coefficient being .564, the corrected .721. The correlation is slightly lower than that for A. It was hoped that counting an unfavorable answer, even though it appeared only once, might improve the reliability of the test; but when the

measure is compared with the comparable ones for A and B separately, we see that this is not the case. The inclusion of changed and unchanged items does not give a more consistent measure of the organization of the individual's responses.

(b) Reliability of same scores. In similar fashion, the reliability, obtained from scoring an item as unfavorable only if it

TABLE 1

ANALYSIS OF THE RELIABILITY OF THE INVENTORY

TEST SCORES	TYPE OF MEASURE	CORRELATION OR RELIABILITY COEFFICIENT	PREDICTABILITY ABOVE CHANCE IN PER CENT [1]
First testing (A)	Half-test reliability	.589	19
Second testing (B)	Half-test reliability	.778	37
Total (T)	Half-test reliability	.564	17
Same scores	Half-test reliability	.643	23
First testing (A)	Test reliability (split-half)	.741	33
Second testing (B)	Test reliability (split-half)	.875	52
Total (T)	Test reliability (split-half)	.721	31
Same scores	Test reliability (split-half)	.783	38
Change scores and difference scores	Correlation	.612	21
First and second testing	Test-retest reliability	.878	52

appeared in both A and B, was calculated. The uncorrected correlation was .643, the corrected .783. This coefficient is still below both the test-retest coefficient between A and B, and the corrected split-half of B. The reliability of the questionnaire, at least in its present form, is not increased by utilizing only the more stable responses. This result may be due in part to the fact that the elimination of the changed items automatically reduces the possibility of variability among the test scores.

(c) Correlation between change scores and difference scores. The correlation between the change scores and the difference scores is lower than any of the uncorrected reliabilities. This finding would indicate that changes in unfavorable responses involve a good deal more than the mere addition of new unfavorable responses to the subject's score, or the deleting of previous unfavorable responses. The mean change score, 4.10, is greater than the mean difference score of 1.98. The critical ratio is 5.58, which, together with the correlation between the difference and

[1] Predictability determined from $1 - \sqrt{1 - r^2}$

change scores of .612, would indicate that the difference and change scores are not mere substitutions for one another. An analysis of the change scores, in other words, will involve more than the ordinary analysis of reliability.

2. *Intercorrelations of the frequencies of A, B, and T and the critical ratios of A, B, and T.* The intercorrelations among the frequencies of unfavorable responses for each item among A, B, and T, and the intercorrelations among the critical ratios of these items are presented in Table 2. The correlations are rather high among the critical ratios and extremely high among the frequencies. However, since A, B, and T differ from each other with respect to other factors (i.e., difference in reliability, differences in mode of determining an unfavorable response), all three variables are retained for the analysis that follows.

TABLE 2

INTERCORRELATIONS OF THE FREQUENCIES OF UNFAVORABLE RESPONSES AND INTERCORRELATIONS OF THE CRITICAL RATIOS OF THE 42 ITEMS THAT ARE CONTAINED IN A, B, AND T

VARIABLES	INTERCORRELATIONS OF THE FREQUENCIES		INTERCORRELATIONS OF THE CRITICAL RATIOS	
	B	T	B	T
A	.982	.993	.847	.925
B	...	.983	...	.858

3. *Intercorrelations among the frequencies, critical ratios, change scores, and same scores for the items of A, B, and T.* For each of A, B, and T, intercorrelations among the frequencies of unfavorable responses, critical ratios, change scores, and same scores were obtained for both the entire 42 items and for 30 items. Correlations utilizing 30 items were obtained in order to reduce the influence of items with low frequencies. Such items were bound to have small critical ratios since unfavorable responses would be scarce. The highest critical ratio of the 12 items eliminated in order to form the group of 30 items was 2.18. For the remaining items, therefore, we could expect a fairly high degree of discriminating capacity. Intercorrelations among the four variables for A, B, and T separately are presented in Table 3. Our chief interest lies in the correlations of the

change scores with the frequencies and the critical ratios, since we wish to see whether those very important features of the items—the frequency with which they appear as symptoms and their discriminating capacity—bear any systematic relationship to the extent to which the subjects' responses are unstable.

TABLE 3

INTERCORRELATIONS AMONG THE FREQUENCIES OF THE UNFAVORABLE RESPONSES, CRITICAL RATIOS, CHANGE SCORES, AND SAME SCORES FOR A, B, AND T FOR 42 AND 30 ITEMS [2]

VARIABLES	1	2	3	4
A				
1. Frequency of unfavorable responses	....	.705	.882	.976
2. Critical ratios	.392	....	.595	.692
3. Change scores	.733	.471	....	.783
4. Same scores	.943	.404	.565	....
B				
1. Frequency of unfavorable responses	....	.710	.851	.984
2. Critical ratios	.576	....	.635	.686
3. Change scores	.704	.450	....	.783
4. Same scores	.966	.558	.565	....
T				
1. Frequency of unfavorable responses	....	.575	.912	.964
2. Critical ratios	.290	....	.532	.549
3. Change scores	.775	.365	....	.783
4. Same scores	.915	.265	.565	....

The correlations with the 42 items are in all cases higher than the corresponding correlations with the 30 items, but the patterns formed by the two sets of correlations are similar. A very significant feature of the correlations is the high relationship of frequency with the other three variables; that is, the extent to which an item is present as a symptom in the general membership of our population seems to be definitely related to the extent to which responses to it are altered, the extent to which unfavorable responses are given at each of the two testings, and the extent to which the items discriminate two extreme groups. There is no contradiction in the fact that the total unfavorable frequency is positively correlated with both the same and change scores. Each represents a measure of frequency of unfavorable responses

[2] The correlations for the 42 items are above the diagonal; those for the 30 items are below the diagonal.

too. The above correlations are, however, partly spurious, since the total frequencies of unfavorable responses are made up of both change and same scores and since, therefore, part of the correlations may be due to the presence of identical elements. The change and same scores do not enter into the total frequencies in equal proportions, as there are on the average about 3.5 same frequencies to 1 change frequency. The correlation between same and change scores alone would represent a correlation between two definitely distinct frequencies of unfavorable responses. These correlations are fairly high: .783 and .565 for 42 and 30 items, respectively.

The correlations with the critical ratios of the various measures of frequency are lower than those that have been found among the frequencies themselves. All of these correlations that have been determined for 42 items, however, are significant. For the 30 items, those determined from the A and B data are significant, while 3 of the correlations in the T groupings are less than four times their probable errors. For the smallest of these, the chances are 97 out of 100 that the correlation is greater than zero. For the 42 items, the correlations of the total frequency with the critical ratios are the highest in all cases (A, B, and T), with the same-score correlations following, although they are not very much higher than the correlations with the change scores. For the 30 items there is no consistent order. The differences between the obtained correlations may very well be due to sampling errors.

On the whole, it would be difficult to conclude from these data which was more significantly related to the critical ratios, the same or change scores. At any rate, differences between correlations for same and change scores, obtained under similar conditions, are not very marked. Since the frequencies are influenced by both same and change scores, it is interesting to note that the frequency correlations are not much above the other correlations for the 42 items, and are highest only once when the 30 items are used. It is irrelevant to say that since the change score represents a frequency, and the frequency is significantly related to the critical ratios, it is not surprising that the change score is related significantly to the critical ratio. As far as the psychological analysis is concerned, it is important to know that those items to which, for some reason, the subject alters his response,

are about as closely related to the discriminating capacity of the items as those items for which there is no such alteration.

4. *Analysis of the changed responses with reference to the nature of the item.* In classifying the items it was desirable to set up dimensions with only a few categories, since the number of questions was so small. We found it possible to set up two main dimensions, each having three categories. All the items could be classified so as to fall into one category of each of the two dimensions.

The first dimension dealt with the subject matter of the item which was classified as (a) factual, (b) relating to physical symptoms, and (c) relating to non-physical factors having to do mainly to attitudes, desires and feelings. Both of the latter two categories were regarded as being subjective. In both cases, judgments of the subject as to whether he had experienced the state or symptom were involved in making the response. The number of items falling into the three categories were 6, 16, and 20, respectively.

The second mode of description involves the frequency factor. Here the 42 items may be classified into three groups depending on whether the word "ever" or "often" was used, or whether the frequency of the symptom was not stated but left to the judgment of the subject. The last grouping was called "indefinite" items. In cases where phrases synonymous with "often" were used, the items were classified in the "often" group. The frequencies of items according to the above classification were 14, 10, and 18, respectively. In other inventories, such as House's revision of the Woodworth Personal Data sheet (**10**), the form of the item is controlled. House begins his inventory with the following statement: "You will assume that each of these statements is prefixed by the remark: This problem has occurred in my life" (p. 53). In classifying the items according to both dimensions the 6 "factual" items have been omitted, making 6 combinations, such as "physical-ever," "physical-often," etc. In some cases, this fractionization results in samplings whose size is of doubtful value. In Table 4, the median change score for each classification is given. We have clearly shown that a relationship exists between the change scores and the frequency of unfavorable responses, so that the change-score medians would

be an incomplete statement of the facts. Accordingly, the ratio between the change scores and the frequency of unfavorable response for A, B, and T of each item has been calculated. The median ratios in each category are listed.

TABLE 4

FREQUENCIES OF CHANGED RESPONSES OF THE VARIOUS CLASSES OF ITEMS

CLASSES OF ITEMS	NUMBER OF ITEMS	MEDIAN NUMBER CHANGED RESPONSES	MEDIAN PER CENT CHANGED RESPONSES	MEDIAN RATIO OF CHANGE SCORE TO UNFAVORABLE RESPONSES IN A	MEDIAN RATIO OF CHANGE SCORE TO UNFAVORABLE RESPONSES IN B	MEDIAN RATIO OF CHANGE SCORE TO UNFAVORABLE RESPONSES IN T
Physical	16	13	8.8	.51	.58	.41
Non-physical (attitudes, etc.)	20	21	14.3	.42	.49	.36
Factual	6	4.5	3.1	.27	.37	.26
"Ever"	14	9	6.1	.38	.43	.33
"Often"	10	11	7.5	.52	.50	.44
Indefinite frequency	18	23	15.6	.46	.51	.39
Physical—"Ever"	4	18.5	12.6	.46	.59	.41
Physical—"Often"	7	4	2.7	.48	.50	.39
Physical—Indefinite	5	22	15.0	.56	.71	.45
Non-physical—"Ever"	4	10.5	7.1	.40	.35	.32
Non-physical—"Often"	3	24	16.3	.58	.70	.48
Non-physical—Indefinite	13	23	15.6	.40	.50	.36

An examination of Table 4 shows that the lowest frequencies of change scores occur for the factual items, the highest for the non-physical items in the subject-matter grouping. The indefinite frequency group is the highest of the frequency classification, with the other two about equal. When the changes are examined with reference to the frequency of the neurotic responses, however, we find that the groupings tend to be closer to each other and that some inversions of the previous orders occur. For the subject-matter dimension, the factual items still remain low, but the physical-symptoms ratio becomes the highest in the group. In the case of the frequencies, though the "ever" questions tend to remain low, the "often" questions come up slightly above the "indefinite" items. The number of items in each category is too few to warrant rigid statistical evaluation. Although the hierarchies obtained, therefore, must be viewed with caution, it should be clear that hierarchies of changed responses reported in previous studies might be very different if the factor of the

frequency of unfavorable response were considered. The hierarchies reported might not represent anything more than the fact that, in the population sampled, some classifications of items were answered in the unfavorable direction more than others were.

The fractionization of the items results in a fair degree of uniformity when the ratios are considered. The absolute change scores vary considerably, indicating that the frequency of unfavorable response is not uniform for the different classifications of items.

The proportion of items that are changed is much smaller than that which has appeared in past reports. In our different categories the range of changes is from 3.1 per cent to 16.3 per cent, as contrasted with average changes of over 15 and 20 per cent. For our subjects, however, the average number of unfavorable responses was only 22.6 per cent. It is from this group of responses, scarcely one-fifth of the total number, that the change scores must come, the maximum number of changes possible being equal to the sum of all unfavorable responses on the two testings.

SUMMARY AND DISCUSSION

In the course of the present study it was found that there were significant correlations between: (1) the changed responses and the frequency of unfavorable responses; (2) the changed responses and those items which were answered in the unfavorable direction at each testing (same scores); (3) the changed responses and the critical ratios between the extremes of the subjects; (4) the frequency of neurotic response and the critical ratio. From these results it was concluded (5) that analyses of the frequency with which responses were changed in answering a psychoneurotic inventory had to take into consideration the frequency with which there were neurotic answers to the items. (6) An analysis of the items showed that the responses were changed most frequently to those items which used the word "often" in connection with a symptom and which made no statement concerning the number of times the symptom had been experienced by the subject.

The meaning of our results depends a good deal upon the

validity of the psychoneurotic inventory that we utilized. The 42 questions assembled by Garrett and Schneck (7) reliably discriminated between a group of Columbia College freshman and a group of abnormals. For the 25 questions that were especially diagnostic, the critical ratio of 12.8 was obtained. Our own subjects had higher neurotic scores than the Columbia freshman, but the critical ratios, when they were compared with the abnormals, were satisfactory. For the first testing, the critical ratio was 2.74, for the second 4.0. With the 25 most diagnostic questions the critical ratio was 8.87 for the first testing and 9.31 for the second. We can at least conclude that the bulk of the questions of the inventory used distinguish our subjects from the abnormals used by Garrett and Schneck.

Assuming that our items are reasonably validated as a whole, we find that those items answered most frequently in the neurotic direction are the ones for which the responses of our subjects are most frequently changed. This finding seems to be somewhat consistent with the neurotic symptom pattern which tends to variability. Those items which are most frequently aspects of the pattern are the ones most frequently interchanged and, therefore, the ones most frequently altered. We cannot eliminate the judgment aspect of the response as shown by changes even in "factual" questions; furthermore, the specific sets of an individual operating at different times may affect his judgments. Responses which in the general population rarely appear as symptoms will, therefore, rarely be changed. A factor making for change might be the variable attitude of the subject towards the symptom, an attitude which would enable him to recognize it at one time and not at another.

According to our results, those symptoms which appear most frequently in our subjects as a whole discriminate our extreme groups most adequately. Since we have found significant differences between our group and the abnormal group, particularly on the special 25 items, it is likely that where an item distinguishes our two extremes, the factor involved is probably not irrelevant (as may occur in cases where the criterion of internal consistency alone is used), but related to the neurotic pattern. With this consideration in mind, it could be maintained as a consequence of our statistical results that those items answered most frequently

in the neurotic direction are the best discriminators. Items to which the responses are altered, moreover, do not thereby lose their discriminating ability.

BIBLIOGRAPHY

1. BAIN, R. Stability in questionnaire response. *Amer. J. Sociol.*, 1931, **37**, 445–453.
2. BENTON, A. L. The interpretation of questionnaire items in a personality schedule. *Arch. Psychol.*, 1935, No. 190.
3. BENTON, A. L., AND STONE, I. R. Consistency of response to personality inventory items as a function of interval between test and retest. *J. soc. Psychol.*, 1937, **8**, 143–146.
4. EDGERTON, H. A., AND PATERSON, D. G. Table of standard errors and probable errors of percentages for varying numbers of cases. *J. appl. Psychol.*, 1926, **10**, 378–391.
5. FARNSWORTH, P. R. A genetic study of the Bernreuter Personality Inventory. *J. genet. Psychol.*, 1938, **52**, 3–13.
6. FRANK, B. Stability of questionnaire response. This JOURNAL, 1936, **30**, 320–324.
7. GARRETT, H. E., AND SCHNECK, M. R. A study of the discriminative value of the Woodworth Personal Data Sheet. *J. gen. Psychol.*, 1928, **1**, 459–471.
8. HARVEY, O. L. Concerning the Thurstone "Personality Schedule." *J. soc. Psychol.*, 1932, **3**, 240–251.
9. HOLLINGWORTH, H. L. *Psychology of the functional neuroses.* New York: D. Appleton, 1920.
10. HOUSE, S. D. A mental hygiene inventory. *Arch. Psychol.*, 1927, No. 88.
11. LANDIS, C. Questionnaires and the study of personality. *J. nerv. ment. Dis.*, 1936, **83**, 125–134.
12. LANDIS, C., AND KATZ, S. E. The validity of certain questions which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934, **18**, 343–356.
13. LENTZ, T. F. Reliability of the opinionaire technique studied intensively by the retest method. *J. soc. Psychol.*, 1934, **5**, 338–364.
14. NEPRASH, J. A. Reliability of questions in the Thurstone personality schedule. *J. soc. Psychol.*, 1936, **7**, 239–244.
15. PAGE, J., LANDIS, C., AND KATZ, S. E. Schizophrenic traits in the functional psychoses and in normal individuals. *Amer. J. Psychiat.*, 1934, **13**, 1213–1225.
16. SMITH, M. A note on stability in questionnaire response. *Amer. J. Sociol.*, 1933, **38**, 713–720.
17. SYMONDS, P. M. *Diagnosing personality and conduct.* New York: Century, 1931.
18. THURSTONE, L. L., AND THURSTONE, T. G. Neurotic inventory. *J. soc. Psychol.*, 1930, **1**, 3–30.
19. WILLOUGHBY, R. R. Some properties of the Thurstone personality schedule and a suggested revision. *J. soc. Psychol.*, 1932, **3**, 401–424.
20. WOODWORTH, R. S. *Personal data sheet.* Chicago: Stoelting and Co.

AGE AND INTELLIGENCE OF A GROUP OF JUVENILE DELINQUENTS

BY CECIL W. MANN
Claremont Colleges, Claremont
AND HELENE POWNER MANN
Juvenile Hall Clinic, Los Angeles

THE findings of the more significant studies during the past few years giving analyses of the ages and intelligence of juvenile delinquents provide the background for the present investigation. In his classic study, Burt (7) found the mean IQ of 200 delinquents in England to be "about 89." Ball (3) found the mean IQ of 146 delinquent boys coming before the Juvenile Court in San Francisco to be 85.4. Snyder (20), surveying a group in the Pennsylvania Training School, found a mean IQ for delinquent boys of 71.25, for girls a mean IQ of 71.46, a difference of means which is not statistically significant. In a survey of 300 girls in the Indiana State School for Girls, Beane (4) found a mean IQ of 82. Of this group, 78 per cent were 14 years and over, while 29 per cent were between the ages of 15 and 16 years.

McClure(15) in Ohio found a mean IQ for 435 delinquent boys of 80.03 with a σ of 12.87, and for 167 girls a mean IQ of 77.27 with a σ of 13.68, giving a difference between means of no statistical significance. Scudder and Beam (19) reporting in 1934 on a study of 14,000 juvenile delinquents in Los Angeles over a period of three years, found 74.72 per cent to be between the ages of 14 and 18 years; of the total group, 51.66 per cent had IQ's of 90 or more, 20.87 per cent IQ's of between 80 and 90 and 27.47 per cent IQ's of less than 80. Lane and Witty (13) in Illinois found a mean IQ for male delinquents of 87.96. In a survey of institutionalized boys in Whittier State School, California, Fenton (8) found a mean IQ of 91.7, but suggested that the relatively high mean was probably due to the policy of the school in refusing to admit feeble-minded boys.

That there are selective factors operating with respect to the

mean IQ of delinquents in institutions is obvious; the policy of one may be such as to inflate the mean, of another to deflate it. It should be remembered, too, that the sample of offenders which reaches the Juvenile Court is a relatively narrow one. At least two factors, intelligence and socio-economic status, tend to select the delinquents who eventually go before the Court. It is not unlikely, for instance, that brighter children—for a number of reasons—will escape detection. Moreover, many who are caught may be returned to their homes if it appears likely that the social status of the home is such that the parents are able and willing to provide adequate control. For these reasons at least, and possibly also for others, it would be unwise to postulate low intelligence as a cause of juvenile delinquency. All that can be inferred is that children of low intelligence coming from areas which are socially underprivileged have a greater chance than most others of appearing before the Juvenile Court.

The present study is concerned with a group of 1,731 delinquent boys and girls who have appeared before the Juvenile Hall Clinic in Los Angeles during the period of two and a half years from March, 1936, to September, 1938. The IQ's were computed from the results gained by administering the New Revision of the Stanford-Binet Scale (Form L), which was introduced to the Clinic in March, 1936. Slightly more than 1,600 of the tests were administered by the two examiners permanently employed at the Clinic, the remainder by one examiner who substituted during the absence of the senior psychological examiner.

AGE ANALYSIS

The distribution of results in age and sex groups is given in Table 1.

Of the total group, 1,061 (61.29 per cent) are boys and 670 (38.71 per cent) are girls; 1,155 (66.73 per cent) of the number are 14 years and over, while 625 (36.10 per cent) are in the 15- and 16-year groups. Of the boys, 678 (63.90 per cent) are 14 years and over, the greatest number, 212 (19.99 per cent), being in the 15-year group. Of the girls, 477 (71.19 per cent) are 14 years and over, the greatest number, 129 (19.25 per cent), being in the 16-year group.

The percentage of delinquent girls 14 years and over is significantly higher than the percentage of boys in the same group, the value of D/σD being 3.16. It is possible that this difference

TABLE 1

COMPARISON OF MEAN IQ (FORM L) OF BOYS, GIRLS AND TOTAL AGE GROUPS

AGE GROUP YEARS	N	MEAN	S.D.	N	MEAN	S.D.	Dm	D/σD
BOYS								
5-8	76	82.92	17.44					
9	35	80.43	16.24					
10	45	83.44	14.96	383	82.45	17.33		
11	52	82.09	16.56					
12	66	82.91	18.19					
13	109	82.27	18.25				3.81±1.13	3.37
14	122	85.77	18.05					
15	212	86.10	18.22					
16	181	86.72	19.00	678	86.26	18.75		
17	113	87.00	19.50					
18	50	84.80	20.03					
	1,061	84.88	18.35					
GIRLS								
5-8	52	83.82	16.10					
9	12	82.83	17.33					
10	21	84.85	16.14	193	82.51	18.79		
11	27	79.96	21.94					
12	32	78.56	19.90					
13	49	84.04	17.39				2.08±1.25	1.66
14	94	85.51	18.84					
15	103	85.15	20.22					
16	129	83.62	17.76	477	84.59	19.09		
17	108	84.91	20.00					
18	43	83.37	19.60					
	670	83.77	19.05					
TOTAL								
5-8	128	83.13	17.20					
9	47	82.04	16.61					
10	66	83.89	15.36	576	82.44	17.72		
11	79	81.36	18.77					
12	98	81.49	18.87					
13	158	82.82	18.02				3.02±0.93	3.26
14	216	85.65	18.29					
15	315	85.79	18.90					
16	310	85.35	18.36	1155	85.46	19.09		
17	221	85.98	19.78					
18	93	82.75	19.95					
	1,731	84.45	18.63					

might be accounted for by the differences in the ranges of offenses between the sexes. The range of delinquencies for boys is much greater than for girls. It includes, among other offenses, theft, larceny, arson, truancy and sex offenses, while for the girls the majority of delinquencies are sex offenses.

At the same time, as will be seen from Table 2, the mean ages of the sex groups vary but little. The difference between the mean ages for the sex groups is 0.4 years, a difference which is not significant, the value of $D/\sigma D$ being 0.27.

TABLE 2

COMPARISON OF MEAN AGES OF SEX GROUPS

SEX	N	MEAN	S.D.	D_m	$\frac{D}{\sigma D}$
Boys	1,061	14.32	2.95	0.40	0.27
Girls	670	14.72	3.00		

It has been suggested on occasions that there is a greater variability in age among male than among female juvenile delinquents. An analysis of variability, shown in Table 3, fails to reveal any significant difference in this respect, the value of $D/\sigma D$ being 0.45.

TABLE 3

COMPARISON OF VARIABILITY OF AGES OF SEX GROUPS

SEX	N	S.D.	D	$\frac{D}{\sigma D}$
Boys	1,061	2.95	0.05	0.45
Girls	670	3.00		

An analysis of the distribution showing percentages of boys and girls at age levels 5 to 18 years indicates that there is no significant difference in the percentages of boys and girls in the various groups, save for the interval containing those of 17 years. At this level the percentage of girls is significantly greater than that of the boys, the value of $D/\sigma D$ being 3.2. Two factors are likely to be operative at this level to bring about this result. In the first place, the majority of delinquencies among girls of this

age are sex offenses, for the discovery of which there is more chance among girls than among boys. Secondly, there is a possibility that the authorities in dealing with delinquents of this age are likely to delay as long as possible sending girls to the adult court, whereas they are less likely to exercise the same amount of sympathy towards boys of the same age.

TABLE 4

COMPARISON OF PERCENTAGES OF BOYS AND GIRLS AT DIFFERENT AGE LEVELS

AGE GROUP	TOTAL		BOYS		GIRLS		D	$\frac{D}{\sigma D}$
	N	PER CENT	N	PER CENT	N	PER CENT		
5–8	128	7.40	76	7.16	52	7.76	0.60	*
9	47	2.73	35	3.30	12	1.79	1.51	*
10	66	3.81	45	4.24	21	3.13	1.11	*
11	79	4.57	52	4.96	27	4.03	0.93	*
12	98	5.67	66	6.22	32	4.78	1.44	*
13	158	9.14	109	10.27	49	7.32	2.95	2.17
14	216	12.49	122	11.49	94	14.04	2.55	1.53
15	315	18.20	212	19.99	103	15.39	4.60	2.41
16	310	17.90	181	17.04	129	19.25	2.21	1.15
17	221	12.72	113	10.62	108	16.12	5.50	3.20
18 & over	93	5.37	50	4.72	43	6.42	1.70	*
	1,731		1,061		670			

* The validity of the formula used for finding the standard error of the percentages is doubtful when the percentages drop below 10. In any case, the values of the differences of the starred percentages are very small for these age groups.

A graphical description of the distribution of the ages of the total group is shown in Figure 1. It will be seen that there is marked skewness towards the higher ages, a skewness which one might expect on logical grounds.

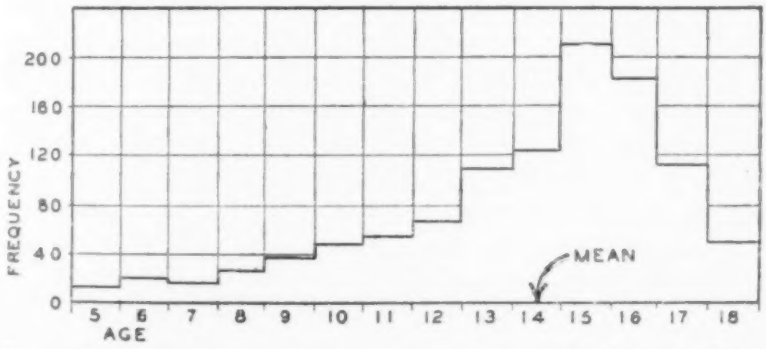


FIG. 1. FREQUENCY DISTRIBUTION OF AGES OF 1,731 DELINQUENT BOYS AND GIRLS

ANALYSIS OF INTELLIGENCE QUOTIENTS

From the results shown in Table 1, it will be seen that the mean IQ of the total group is 84.45, with a S.D. of 18.63. In a comparison of the delinquent group with the group on which the New Revision was standardized, it is evident that the mean IQ of the former is lower and the S.D. higher, the mean IQ of the standardized group clustering about 100 and the S.D. about 16.4 (22).

TABLE 5

COMPARISON OF MEAN IQ (FORM L) OF SEX GROUP

SEX	N	MEAN	S.D.	D_m	$\frac{D}{\sigma D}$
Boys	1,061	84.88	18.34	1.11	1.19
Girls	670	83.77	19.05		

The mean of the group under survey is higher than that given by Snyder (20) and lower than that given by Fenton (8). It is likely that this group is relatively free of the selective factors which operate over the institutionalized delinquents. Many of

TABLE 6

COMPARISON OF VARIABILITY OF IQ (FORM L) OF SEX GROUPS

SEX	N	S.D.	D	$\frac{D}{\sigma D}$
Boys	1,061	18.34	0.71	1.10
Girls	670	19.05		

the delinquents of the present study will be returned to their homes; and, in any case, the test results have been obtained prior to the operation of the selective factors in the institution.

Comparing the mean IQ's of the sex groups, we find, in Table 5, that there is no significant difference between these means, the value of $D/\sigma D$ being 1.19.

A distribution of the percentages of boys, girls, and of the total group in the different IQ levels is given in Table 7. It will be seen that there is not, in any group, a statistically significant difference in the percentages of boys and girls at the various IQ

levels. Nor do we find any statistically significant difference in variability. There is a difference between the S.D.'s of the boys and the girls of 0.71, giving $D/\sigma D$ a value of 1.10.

TABLE 7

COMPARISON OF PERCENTAGES OF BOYS AND GIRLS AT DIFFERENT IQ LEVELS *

IQ	TOTAL		BOYS		GIRLS		$\frac{D}{\sigma D}$
	N	PER CENT	N	PER CENT	N	PER CENT	
130	16	.92	11	1.03	5	.75	
120	56	3.23	33	3.11	23	3.42	.35
110	105	6.06	65	6.13	40	5.97	.13
100	188	10.86	120	11.32	68	10.15	.77
90	299	17.27	189	17.81	110	16.41	.76
80	346	19.99	200	18.86	146	21.79	1.46
70	327	18.89	204	19.20	123	18.36	.45
60	237	13.74	159	14.99	78	11.62	2.03
50	130	7.51	70	6.60	60	8.96	1.80
40	27	1.56	10	.93	17	2.54	
	1,731		1,061		670		

* The mean IQ and the S.D.'s were calculated on a distribution having an interval of 5 IQ points.

A graphical description of the distributions of IQ's is shown in Figure 2. It will be seen that the distribution approximates a normal curve, with a mean round 85.

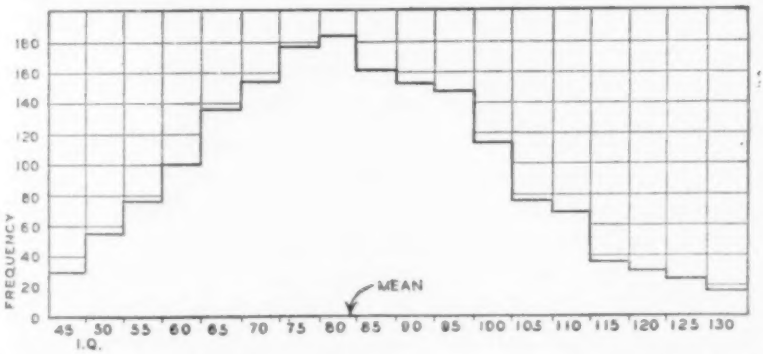


FIG. 2. FREQUENCY DISTRIBUTION OF IQ'S OF 1,731 DELINQUENT BOYS AND GIRLS

Referring again to Table 1, the reader will notice that there is a slight but significant difference in the means of the IQ's of the group of boys 14 years and over and the group under 14, the former having the higher mean. This difference is absent in the results of the girls. It is possible that there is more

control exercised over girls than over boys at the higher ages, particularly in the higher social levels; and this circumstance, coupled with the fact that most of the female delinquencies are sex offenses, may account for the absence of the difference in the distribution of the results of the girls. If, however, we compare the difference of the differences (3.81 and 2.08), we get a value of 1.73, a difference which is not significant, the value of $D/\sigma D$ being 1.03. It is suggested, therefore, that the difference may be due to sampling.

SUMMARY OF RESULTS

1. Of the total group of 1,731 juvenile delinquents included in this study, nearly two-thirds are boys.

2. The mean age of the total group is 14.47 years. The mean age of the boys is 14.32, of the girls 14.72, a difference of means of 0.40 years, which is not significant.

3. The mean IQ of the total group is 84.45. That of the boys being 84.88, of the girls 83.77. There is a difference between these means of 1.11, a difference which is not significant.

4. There are no sex differences of statistical significance with respect to:

(a) mean age
(b) variability of age
(c) percentages in the various age groups
(d) mean IQ (measured on Form L)
(e) variability of IQ (measured on Form L)
(f) percentages in the various IQ levels

5. The group of boys 14 years and over tends to have a slightly higher mean IQ than that of the younger group below 14 years. This difference does not appear in the distribution of the results of the girls.

DISCUSSION

It must be clearly emphasized that any conclusions drawn from the results given in this study are subject to the factors which have operated in the selection of the sample. Although it is the policy of the Juvenile Hall Clinic to administer psychological examinations to all who come before it, at the same time it is obvious that factors outside the influence of the Clinic tend to

select its subjects. It has already been pointed out that factors of intelligence and socio-economic status tend to select the delinquents that are brought before the Juvenile Court. It is doubtful if at any time we can be certain of having a truly representative sample which could be labeled "juvenile delinquent."

With this in mind, we must exercise due caution when we assume that the mean age of delinquents is 14½ years or that the mean IQ is 85 or any other such number. All that can be stated with certainty are the statistics for the sample under consideration. It would hardly have been necessary to emphasize this apparently trite warning but for the common generalizations which are made from similar studies relative to mean ages and IQ's of delinquents. The implications of such generalizations—for instance, that delinquents "generally" have an IQ below 90, and that those whose IQ is above are rare—are of doubtful value. A closer approximation to a general rule is that delinquents having an IQ below 90—because of low intelligence, because of the area from which they come, or for some other reason or reasons—are more likely to be caught in their delinquencies than those whose IQ is higher.

It has already been suggested that nothing in this study can be interpreted to prove that low intelligence is a cause of juvenile delinquency. To do so is to oversimplify the situation. Delinquency can hardly be postulated as an effect following upon any single cause. Intelligence and age, as well as socio-economic factors, are important in contributing to that complex which expresses itself in delinquent behavior.

It is significant, however, that the group which, because of low intelligence, finds greatest difficulty in manipulating an academic curriculum contains the greatest number of individuals who are likely to appear before the Juvenile Court. The members of the borderline group will, at all times and at all school levels, find adjustment to a normal school-curriculum difficult. All but the most daring will, up to the age of 13 or 14 years, accept with some docility the irksomeness of school routine, but more and more will escape at higher age levels even though it be to end up in Court.

That there are no sex differences with respect to mean age and mean IQ is what might have been expected on logical

grounds. For the delinquent, boy or girl, the offense is, at the time, an escape from an unsatisfactory system of behavior. It is apparent that there is for girls, even as for boys, an equal need for escape. The chief sex differences—a problem beyond the scope of this study—would probably be found in the types of escapes employed.

BIBLIOGRAPHY

1. AICHORN, A. *Wayward youth.* New York: Viking Press, 1935.
2. ALEXANDER, F., AND HEALY, W. *Roots of crime.* New York: Knopf, 1935.
3. BALL, R. J. A survey of 146 committed delinquents in San Francisco. *J. juv. Res.*, 1928, **12**, 241–243.
4. BEANE, J. C. A survey of 300 delinquent girls. *J. juv. Res.*, 1931, **15**, 198–208.
5. BRIDGES, J. W. A study of a group of delinquent girls. *Ped. Sem.*, 1927, 34, 187–204.
6. BRIDGES, J. W., AND BRIDGES, K. M. A psychological study of delinquents by group methods. *Genet. Psychol. Monogr.*, 1926, **1**, 411–506.
7. BURT, C. *The young delinquent.* New York: Appleton, 1925.
8. FENTON, N. *The delinquent boy and the correctional school.* Claremont: Claremont Coll. Press, 1935.
9. GLUECK, E. T. Mental retardation and juvenile delinquency. *Ment. Hyg.*, 1935, **19**, 549–572.
10. HALES, W. M. Result of a group intelligence re-test with a reformatory group. *J. juv. Res.*, 1937, **21**, 181–188.
11. HEALY, W. *The individual delinquent.* Boston: Little, Brown, 1927.
12. HEALY, W., AND BRONNER, A. *New light on delinquency and its treatment.* New Haven: Yale Univ. Press, 1936.
13. LANE, H. A., AND WITTY, P. A. The mental ability of delinquent boys. *J. juv. Res.*, 1935, **19**, 1–12.
14. *Juvenile Hall Annual Reports.* Los Angeles: Probation Committee.
15. MCCLURE, W. E. Intelligence of six hundred juvenile delinquents. *J. juv. Res.*, 1933, **17**, 35–43.
16. MATHEWS, J. A survey of three hundred forty-one delinquent girls in California. *J. Delinq.*, 1922, **8**, 196–321.
17. MERRILL, M. A. Mental differences among juvenile delinquents. *J. Delinq.*, 1926, **10**, 415–427.
18. OWEN, M. B. The intelligence of the institutionalized juvenile delinquent. *J. juv. Res.*, 1937, **21**, 199–206.
19. SCUDDER, K. J., AND BEAM, K. S. *Who is delinquent?* Los Angeles: Rotary Club of Los Angeles, 1936.
20. SNYDER, M. A. A comparison of mental traits and attitudes of delinquent boys and girls. *J. juv. Res.*, 1931, **15**, 181–191.
21. STONE, C. P. A comparative study of the intelligence of 199 inmates in the Indiana Reformatory. *J. crim. Law Criminol.*, 1921, **12**, 238–256.
22. TERMAN, L. M., AND MERRILL, M. A. *Measuring intelligence.* New York: Houghton Mifflin, 1937.
23. TOLMAN, R. S. Differences between two groups of adult criminals. *Genet. Psychol. Monogr.*, 1938, **20**, 353–458.

DEVELOPMENTAL CHARACTERISTICS AND PERSONALITIES OF CHRONIC ALCOHOLICS *

BY MARY PHYLLIS WITTMAN

Elgin State Hospital
Elgin, Illinois

MANY psychiatric and psychoanalytic studies have been made recently to determine the type of developmental history characteristic of the chronic alcoholic. Prominent among the reports are those of Wall (**14**), Knight (**8**), Tillotson and Fleming (**13**), and Strecker and Chambers (**12**).

The majority of these studies have stressed the same type of etiologic developmental factors as characteristic of the chronic alcoholic's history. These factors include a doting, over-solicitous mother and a comparatively stern, forbidding father, the latter a person who inspired respect and awe or fear, and who displayed inconsistent tendencies of severity and indulgence, thus producing in the child a feeling of insecurity and helpless dependence. The personality traits thought to be characteristic of the chronic alcoholic include narcissism, emotional immaturity or instability, infantilism, passivity and dependence, oral eroticism and latent homosexual trends, often with compensating Don Juan tendencies.

In the studies upon which the above conclusions have been based there has been no specific statement of the technique used in gathering the data, nor of any fixed criteria by which the data were compared and evaluated. The conclusions have been based, evidently, upon the examiner's interpretations of the histories of the patients and the information gained from the patients themselves. Although this procedure is, of course, the usual psychiatric and psychoanalytic one, it would seem that a

* This study, undertaken at the suggestion of Dr. Charles F. Read, Managing Officer, Elgin State Hospital, Elgin, Illinois, owes much to his advice as well as to the counsel and encouragement of Dr. Andrew Brown, Institute for Juvenile Research, Chicago, Illinois. I wish also to acknowledge the courtesy of Dr. F. Gerty, Superintendent of Cook County Psychopathic Hospital, in allowing me to secure subjects among the patients in his institution. A part of this paper was read at the American Psychological Association meeting, September 1938, Columbus, Ohio.

more objective approach might now be used in testing these hypotheses. J. O. Chassell (4), for example, has constructed a scale suggestive of a method to be used in an objective approach to the study of those determinants of personality formation that have been stressed by various etiologic theories of chronic alcoholism.

Following his lead, one could devise a scale that would include all the points thought to be characteristic of the developmental etiology of the so-called "chronic alcoholic personality." The claims of various writers might then be evaluated in terms of this objective scale. For example, the claim that mothers of alcoholics are on the whole over-solicitous would no longer depend upon the examiner's subjective interpretation of what the patient and his social-service history say directly or indirectly about this point, but rather upon the patient's answer to a standardized test situation. This procedure would seem justifiable, since the true degree of solicitude shown by the mother—which, after all, is a relative point—is not so important or significant as the subject's own estimation and interpretation of her solicitude.

The validity of such an approach may be questioned, of course, since it is admittedly difficult for individuals to classify themselves into rigidly defined trait levels. Chassell (4) very aptly answers the criticism by stating, "Exponents of the heterogeneity of human personality may point out that no tabulated record can give the true portrayal of an individual's unique self. Indeed, the alternatives under the various headings seldom fit precisely. Granted! Yet generalizations have been made and will be made regarding types of social situation and types of personal response. If such generalizations are justified, then the attempt to isolate and characterize these types as separate alternatives must also be justified."

PROCEDURE

The experimental subjects for this study included 100 patients from the Elgin State Hospital and Cook County Psychopathic Hospital with a diagnosis of chronic alcoholism, without psychosis.[1] The control subjects were 100 volunteers, equated as

[1] Male subjects alone were used because of the limited number of women diagnosed as chronic alcoholics available, and since the addition of these few would have required the evaluation of still another uncontrolled variable, i.e., sex difference.

to age, education and nationality with the alcoholic group. Only subjects between 25 and 50 years of age with at least graduation from English-speaking grammar schools were included, since the test material required a fair degree of reading ability.

A scale incorporating part of the material from the Chassell scale, together with added topics suggested by the various theories as to the personality development and traits of the chronic alcoholic, was used. The complete scale is made up of fifteen large divisions:

I. Sociological Data
II. Mother Relationship
III. Father Relationship
IV. Sibling Relationship
V. Home Life
VI. Religion and Standards
VII. Sex Development
VIII. Love Affairs
IX. Physical Development
X. Intellectual Develpoment
XI. Vocational Adjustment
XII. Social Adjustment
XIII. Emotional Adjustment
XIV. Marital Adjustment
XV. Miscellaneous

Each one of the large headings has from 2 to 20 subheads, which in turn are, with the exception of I, divided into four graduated descriptive levels taken from or modeled after the ones in Chassell's Inventory and labeled a, b, c, and d. As an example, one of the subheads under Mother Relationship is given here.

MOTHER'S TENDENCY TO DOMINEER

a. Demands that her own plans and ideas be carried out. Strict obedience—Mother knows best.
b. Likely to express grief, disappointment, dissatisfaction at deviations from her wishes.
c. Never insists on her way alone—makes suggestions.
d. Cooperates in subject's decisions, even when not fully understanding or agreeing with them. Encourages him to make his own decisions.

The large divisions II through V refer directly to the childhood and early adolescent periods of the subject, but for the other divisions no time limits are given.

The alcoholic subjects[2] and most of the control group were given the history scale individually and in the presence of the examiner.[3] As a routine procedure the examiner read aloud from a second copy of the scale the first two or three headings and the division levels underneath, explaining the scale and the way it was to be answered. The patient, who sat on the other side of the desk from the examiner, simply put down the letter standing for a certain level, under each number, without being directly observed. This method was purposely adopted to avoid any distaste or embarrassment the subject might feel in indicating a certain descriptive level directly under the gaze of the examiner. In addition, because of the rather personal nature of some of the questions, the subject was asked *not* to put his name on the paper. With both alcoholic and control subjects it was stressed that the study had to do with *group* differences and not individual ones, that the subject's answers were not to be associated with his name at all, but simply with those of the rest of the group. In this way we hoped to get a consciously honest answer, at least as far as the subject's own insight was concerned, and to rely upon inferences and generalizing about subconscious motivations only in a general discussion of the results.

The data collected from this scale were handled in two ways:

(1) A numerical value was assigned to each of the four letter levels and in this way the place on a scale continuum determined for the specific developmental factors. Thus, the differences between the alcoholic and the control group would indicate which traits are more or less characteristic of the chronic alcoholic.

(2) The scale (Part II through Part XV) was submitted, individually, to a group consisting of ten psychiatrists, two psychoanalysts and three psychologists.[4] They were asked to rate the four levels under each variable—from the one most likely to contribute toward general adjustment at the top, down to the

[2] The alcoholic subjects were friendly and cooperative in their reactions. Only one took advantage of the offer to withdraw, stating rather indignantly that he was "in the hospital for treatment, not to act as a human guinea pig."

[3] The only exception to this rule was in the case of some staff physicians and internes used among the controls, who undoubtedly understood the scale itself and all the terms used, so that they were able to answer the scale by themselves and return it to the examiner.

[4] Part I dealt with sociological and statistical data and was not divided into graduated levels that could be weighted.

one most likely to lead toward maladjustment and lack of stability at the bottom. The four graduated levels under each variable were arbitrarily assigned weight values from zero at the lowest level up to three at the highest level. These weights were then used in re-scoring the self-ratings of the chronic alcoholic and control groups and thus securing "adjustment" scores for each trait.

RESULTS

Sociological Data. The alcoholics as a group came from somewhat larger families than did the controls (2.7 siblings for the experimental average, as compared to 1.3 for the controls). Their relative position among the siblings, however, was approximately the same: 1.5 for alcoholics, 1.2 for controls. This fact does not corroborate the frequently expressed picture of the chronic alcoholic as usually the "spoiled baby," that is, the youngest child or only child, pampered since birth and hence conditioned to demand and take any desired thing, no matter what the consequences.

The average age at which the alcoholic first used liquor (not just tasted it) shows a decided difference from the control age. The average for the alcoholic group was 20 years, and for the control group 26 years. The indication is, therefore, that the chronic alcoholics as a group start drinking at a significantly earlier age than the controls. This point has been stressed by many other studies. Evidently, then, the early as well as the continued use of alcohol is a significant factor in the development of chronic alcoholism.

The economic status of the alcoholic group was determined from the statistical data sheet on file in each patient's folder at the Elgin State Hospital, while for the controls their own estimates of status were accepted. The difference between the number in the dependent level for the two groups (27 per cent of the alcoholics, 0 per cent of the controls) might be looked upon as a result rather than a contributory cause of the chronic alcoholism. Sociologists have noted the apparently higher incidence of certain social maladjustments somewhat allied to chronic alcoholism (e.g., delinquency, crime, prostitution, etc.) among the lower socio-economic levels. The socio-economic

status itself might be considered a contributing factor to the development of chronic alcoholism in the same way that it constitutes a possible factor in delinquency and crime.[5]

The marital status of the two groups was also compared. A larger percentage of the controls were single, but also a larger number were married, at the time of the examination, than in the alcoholic group. The fact that thirty per cent of the alcoholics were separated might be considered either a result of the alcoholism or a cause of it. Twenty-one per cent of the experimental group had been married more than once, compared to only six per cent of the controls.

Mother Relationship. The topics included under this heading stress the relationship between the subject and his mother during his childhood and adolescent years, while he was still under the family roof. In those dealing with the mother as an individual, no decided difference is found between the two groups, although the typical mother of the alcoholic is somewhat more of a "home body" and slightly less stable emotionally—that is, in her son's opinion, of course.

The mother's interest and personal concern for the subject, though pronounced, was approximately the same for the two groups. This finding does not bear out the psychoanalytic interpretation of the alcoholic's mother as a person differentiatingly over-protective and solicitous. Such an interpretation is not necessarily disproved, however, since it is quite possible that a mother of this type might not be pictured by her son in those terms. This possibility was suggested by the finding that the alcoholic subjects considered their mothers to show a much stronger tendency to domineer and to express dissatisfaction and disappointment at deviations from her wishes than did the controls. Since the alcoholic *does* rate his mother as definitely more domineering, it is possible that what to the observation of a third person (i.e., the psychiatrist) appears to be over-solicitude and concern, the individual in question interprets as a tendency on his mother's part to think that "mother knows best" and to insist on her plans and ideas for him.

[5] It is admittedly difficult to determine the relative importance of the socio-economic level in delinquency and crime. Analogously, in chronic alcoholism, wealth and position act as a protector and discriminator as far as arrest and commitment to a state or county institution is concerned.

The mother's severity in dealing with her son was approximately the same for the two groups; in other words, it was rated between the two levels (b) "occasional spankings," and (c) "scolds, but quick to make amends." Most of the alcoholic and control mothers alike were placed in (b) "steady general affection—consistent but not marked," although twelve of the alcoholic group checked "very demonstrative, marked exhibition of affection" as against only one of the controls.

The experimental subject indicated a somewhat stronger tendency to sacrifice his own wishes and plans for his mother. This difference may be associated with the alcoholic's rating of his mother as domineering and dissatisfied with deviations from her wishes, or it may be his rationalization as a devoted son, a martyr to mother's wishes.

Idealization of mother was also rated higher by the alcoholic than by the control. A study of responses for these last two items shows a greater tendency for the alcoholic group to answer in a conventionally sentimentalized direction. As compared with the control, the alcoholic (in his own opinion) idolizes his mother and makes a greater sacrifice of his own wishes for her. These responses may be connected with the "narcissism" that is thought by some to be a characteristic of the group.

The adjustment scores under mother relationship are, in every case, lower for the alcoholics than for the controls. The two experience variables where the greatest differences occur are "Mother's tendency to domineer" and "Subject's idealization of mother." These are the two same variables that showed the greatest differentiation in the average levels indicated by the two groups.

Father Relationship. The average father of both the alcoholic and control group was rated as "somewhat ambitious but not more interested in succeeding than in making friends," and as "not outstanding in social success but well liked, with a good reputation." Six per cent of the alcoholics, however, rated their fathers as "rather crude, looked down upon as a crank or ne'er-do-well," a characterization made by none of the controls.

Both the average alcoholic and the control considered their fathers to be "usually well-adjusted but at times upset, irritated

or out of sorts." The typical father for the two groups was "not very demonstrative, but natural and affectionate"; he "showed considerable interest and attention but left bringing up of the subject mostly to the mother."

Most of the subjects in both groups thought that they "understood their father" and were "tolerant and sympathetic of his faults." They also "talked some things over with him but did not have many very intimate and confiding discussions." The only two topics that showed much difference for the two groups were "Subject's fear of father" and "Father's tendency to domineer." The alcoholics, on the whole, considered their fathers to rate between (a) "autocratic, forced obedience, insisted that his plans be carried out," and (b) "tactful but insistent on obedience." The average father for the controls, on the other hand, was rated between (b) and (c) "father gave few commands but many suggestions and advice." The alcoholics rated themselves as "somewhat afraid of father, usually obeyed because of fear not to," while the controls had "no fear, able to differ from father if considered it important enough."

The average father of the experimental group "indulged in alcoholic drinks but not excessively," while the control's father drinks somewhat less. Only one of the controls rated his father as a drunkard as against twelve such ratings among the alcoholics. The alcoholics tended to rate their fathers at the extremes, either "teetotalers" or "habitual drinkers," while the controls placed theirs in the middle range.

The adjustment scores for the variables under father relationship show no such marked differences as those under mother relationship. The two traits in which the greatest relative differences appear are "Father's tendency to domineer" and "Subject's fear of father." Again, these are the two variables showing the greatest difference in average level for the two groups, an agreement which necessarily follows since the letter levels were left in the same relative position when rated as to adjustive level.

This lack of adjustment on the alcoholic's part is, of course, based on his own self-ratings and, although some maladjustment is indicated, the possible mechanisms which might motivate his interpretation of these relationships are not evaluated. For instance, the suggested maladjustment of the alcoholic as related

to his father's tendency to domineer might be due, basically, to his resentment of *any* restraint, social as well as parental. Similarly, his fear as a child of his father might not be due to the father's characteristics, but rather to the expression of a strong feeling of personal timidity on the part of the subject.

Sibling Relationship. The alcoholics rated themselves as somewhat more talented than their siblings, but so too did the controls. As for popularity within the family and amount of companionship and intimacy, both alcoholic and control groups had an average or middle rating. Although both groups rated themselves as slightly more talented than their brothers or sisters, neither considered that there was much jealousy or competition among the siblings, nor much friction among the children. The only factor rated definitely higher by the alcoholics was the tendency of parents to favor the subject above the other children. The average alcoholic rating was near level (c) "comparisons seldom made but a little partiality shown in favor of subject," while the average control considered (b) "subject about on a par with the others."

Home Life. Both of the groups, finding very little overt conflict between their parents, put the average around level (c) "some squabbles, but not severe." The obedience demanded by parents was greater and the amount of freedom and latitude less as estimated by the alcoholics, although the difference is not significantly large. Their response to parental measures of discipline was between (b) "inwardly resentful, sometimes openly defiant" and (c) "docile to avoid trouble"; and for the controls between (c) and (d) "docile, no feeling of rebellion."

An interesting and unexpected finding was the difference in age at which the two groups left home. The average for the controls was 18.8 years, for the alcoholics 23 years. Among the alcoholics there were several who had never left home—that is, they had been still living in the family home when sent to the Cook County Psychopathic or Elgin State Hospital. All the controls, on the contrary, had left home before the initiation of this investigation. This finding may be linked with the relatively large number of alcoholics (27 per cent) listed as "dependent," since many of this group were possibly dependent on their parents and hence living at home.

The adjustive importance of the home-life factors is approximately the same for both groups, except for "Amount of freedom and latitude allowed subject by parents." Even this factor, however, is only relatively differentiating.

Religion and Standards. The attitude of the alcoholic's parents toward religion was near level (a) "sincerely religious; religion important factor in their lives," while the controls rated their parents' attitude nearer (b) "passive interest of average church-member type." The same relative positions were rated in regard to parent's tendency toward religious conservatism as well as parent's tendency to become emotional about differing points of view (moral, religious, political).

The alcoholic's *own* attitude toward religion (church) was placed at level (b) "definitely interested in religion." The controls on the average rated nearer level (c) "interested in church only as a social meeting place." On this attitude the controls showed a much wider range than the alcoholics, a few rating (a) "life centered in religion," but still more counter-balancing with the opposite extreme (d) "never go to church or only when must."

The alcoholic subject's feeling of need for religious security was rated near (a) "strong and insistent desire for feeling of security and peaceful sense of presence of God," that of the controls between level (b) "some need to work out a well-ordered philosophy of life" and (c) "scarcely gives the matter a moment's thought." It was noted that though fourteen per cent of the controls checked level (d) "no feeling of need for religious security," not a single one of the alcoholics did so. Inasmuch as the lack of a feeling of security (security in general, not necessarily religious) is thought by many psychiatrists and psychoanalysts to be fundamental in the etiology of alcoholism, this item is a significant one.

The alcoholics also indicated a stronger sense of sin and guilt, the average falling between level (a) "marked feeling of guilt—done things very wrong that he hates to acknowledge even to himself" and (b) "considerable feeling of guilt—done things he would have difficulty in acknowledging to others." The control average, on the other hand, was at level (c) "occasional qualms of conscience."

Sex Development. The parent's attitude toward sex in general and the respectability of the subject's sources of sex information was approximately the same for both groups. The parental attitude toward the acquisition of sex information by the subjects also showed little differentiation.

The satisfactoriness of present sex adjustment, however, was between "very satisfactory" and "as satisfactory as age and circumstances permit" for the controls, while for the alcoholics the average rating was nearer "at times upset by practices and dissatisfactions."

Love Affairs. The frequency of crushes (strong emotional attachments) on persons of the same sex was rated very low for both groups, with the great majority indicating that they had "never had any." This item had been included because of the often-expressed concept (8) of the chronic alcoholic as a latent homosexual, at either a conscious or unconscious level, one who drinks to overcome the tensions produced by his consciously unacceptable tendencies and fears.

Although the alcoholic and control groups disclaimed any strong emotional attachment for other men, it is extremely doubtful, because of the pronounced social disfavor, if the lack of affirmative answers is significant. Even if conscious and overt homosexual tendencies were present, neither the alcoholic nor the control might be willing to admit them dispassionately; or, on the other hand, the tendencies might not be consciously accepted by the individual.

Interest in the opposite sex and the frequency of love affairs with the opposite sex were both rated significantly higher by the alcoholic than by the control group. This finding agrees with the reports of many investigators who have stressed the Don Juan tendencies of the typical alcoholic.

Physical Development. Under physical fitness the alcoholic group rated themselves lower than the control group, 18 out of the 100 subjects classifying themselves as "delicate, required to pay very close attention to their health." Yet, on the other hand, the alcoholic group rated themselves several points better, although not significantly higher, on ability in physical sports. One would expect a closer relationship between degree of

physical fitness and ability in physical sports than the alcoholics indicated.

In the accounts given by chronic alcoholics, both in mental examinations and in the present scale, there is a seeming tendency to project the causes for their difficulties. Taking this fact into consideration, one might look upon their low ratings in physical fitness as a form of defensive rationalization. Although lowered physical efficiency might be a result of the excessive alcoholic indulgences of the group, this circumstance would not explain the self-ratings of these individuals as possessing greater ability in physical sports than the controls. With this rating in mind, one suspects among the alcoholics a tendency toward self-idealization lacking in the apparently harsher, more realistic self-conceptions of the controls.

Intellectual Development. The ratings of the alcoholic and control groups did not show significant differences as to their intellectual accomplishments in school, nor in their parents' attitude toward their school work. The only marked difference in adjustment average under physical and intellectual development was in regard to "Physical fitness," in which the alcoholics, rating themselves lower as a group than did the controls, have a lower adjustment score.

Vocational Adjustment. Although both groups showed essentially the same degree of ambition, the controls rated themselves significantly higher in success at realizing their vocational objectives. This state of affairs may, of course, be at least partially explained as either a concomitant or as a result of the alcoholic's indulgence.

Social Adjustment. The status of the family in the community was rated significantly lower by the alcoholic than by the control group. The alcoholics, even so, considered themselves to rate somewhat higher in social treatment from companions and associates than the controls did. This tendency of the chronic alcoholic to be a good mixer and socially congenial has been pointed out by many investigators.

The alcoholics considered themselves more sensitive to the opinion of others, however, than did the control group—this in contradistinction to the fact that they are individuals who drink to such excess that they are viewed in a disparaging light by the general public. The rating here may be associated with

the similar self-rating in physical fitness; in both the alcoholic rates himself as sensitive (physically and mentally) and thus less able to withstand hardships and adversities.

In very decided contrast to the lack of freedom from hyper-sensitiveness noted by the alcoholics was the definite freedom from hyper-self-consciousness which they expressed. Between their average rating and that of the control group there was a marked difference. Ordinarily one expects hyper-sensitivity and self-consciousness to co-exist in a given individual, but apparently the alcoholic repudiates any such relationship, so far as he is concerned. This lack of association between sensitivity and self-consciousness is somewhat analogous to the lack of agreement between physical condition and athletic prowess. The explanation may be the same in both cases, i.e., a tendency to idealize and view defensively, with but little if any insight, his own traits and behavior reactions.

Neither of the groups considered themselves to be domineering or to exercise strict authority, nor was there any significant difference between the two groups. This trait rating had been included because Knight's belief (8) that the alcoholic is generally a passive, dependent, and submissive type of individual suggested that the experimental group might rate the lower end of this trait significantly. It must be kept in mind, of course, that these are self-ratings and that quite possibly the chronic alcoholic does not recognize or accept the presence of these traits within himself.

Both groups rated themselves as possessing close, understanding friendships. Knight had found that the chronic alcoholic has many fair-weather friends, but very few, if any, close long-time friends.

The alcoholic group rated themselves somewhat higher as to success in getting along with others than the control group did, although the difference was not especially significant. This success in getting along with people is probably to be associated with their lack of hyper-self-consciousness and their friendly and likeable social traits. Knight found the alcoholics in his study to be passive and dependent in their relationships, yet very personable, friendly, and likeable. He suggested that the chronic alcoholic may, because of a sense of dependence and a fear of inferiority, unconsciously develop a technique for making himself

liked and for developing into a good talker—a "life-of-the-party" type of individual.

General Emotional Adjustment. The alcoholics, obviously enough, rated themselves much lower as to their satisfaction and happiness in life than did the controls. As to the customary level of emotional tension, the average alcoholic rated himself as "somewhat high-strung, often keyed up, works under high nervous tension." The control, on the other hand, considered himself to be "usually calm although upset at times with appropriate provocation."

The majority of the alcoholics, as well as the controls, considered their method of dealing with difficulties and weaknesses as "an adjustive one—faces situations objectively; seeks the best counsel available." This assertion suggests a definite "blind spot" on the part of the alcoholic, since many of these patients, instead of seeking counsel and hence coming to the institution voluntarily, had been committed—even, at times, against their expressed wishes.

Marital Adjustment. The ninety-three alcoholics and the seventy-nine controls who were or had been married rated the factors listed under marital adjustment.

The controls rated their married life as more congenial than the alcoholics. Because of the large number of divorced and separated among the alcoholics, this estimate would probably be expected. The other significant differences in ratings on marital adjustment included the wife's attitude toward drinking and the wife's own drinking habits. The alcoholics rated their wives and ex-wives as less tolerant toward drinking and as less inclined to take any alcoholic drink than did the controls.

These differences between the ratings for the two groups may indicate an actual differentiation, or they may be due to a difference in the relative evaluation of terms by the two groups. The alcoholic's connotation of the terms "liberal," "intolerant," "teetotal," "excessive," etc., as applied to something in which he indulges, is quite possibly in contrast to the control's interpretation.

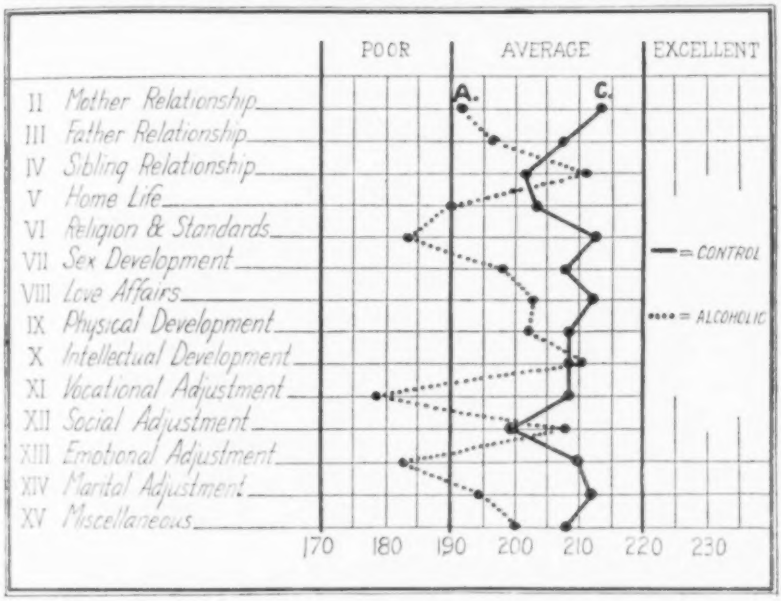
In marital adjustment the trait heading showing the greatest deviation between experimental and control scores was "Congeniality of married life." The alcoholics had indicated definitely

less congeniality than the controls and hence received a much lower adjustment score.

Miscellaneous. Among the trait headings grouped together under "Miscellaneous" the only one that strikingly differentiated the two groups was "Love for parents." The majority of alcoholics indicated a greater love for their mother, while most of the controls had no preference. The latter rating was given the higher adjustment weight.

The accompanying graph gives the average adjustment scores for the different parts of the scales.[6]

ADJUSTMENT SCORE AVERAGES FOR PARTS II THROUGH XV



The divisions showing the greatest differentiation are Vocational Adjustment, Emotional Adjustment, and Religion and Standards. It is possibly significant and certainly an interesting point that all of these divisions have to do with adjustments that the individual makes with his inner self as opposed to external adjustments between the individual and people or things in the environment.

[6] The score range between 190 and 220 points was marked off as average because under the part headings all the adjustment scores for the controls fell within this range.

The divisions, taken as a whole, in which the alcoholics fall definitely below "average adjustment" are again X, XIII, and V.

On three of the divisions, Sibling Relationship, Intellectual Development, and Social Adjustment, the score averages for the experimental group exceeds (although only slightly) those of the control group.

The lowest adjustment average made by the control group was in Social Adjustment, while the alcoholics, on the other hand, had their lowest adjustment scores in Vocational and Emotional Adjustments. The highest division score made by the experimental group was in Sibling Relationship, the next highest in Intellectual and Social Adjustment. The average total adjustment on the scale for the alcoholic group is 195 as compared to 208 for the control group. Both of these scores fall within the range taken as average.

CONCLUSIONS

Claims concerning the presence of certain etiologic developmental factors, stressed by recent psychiatric and psychoanalytic studies, have in general, but not entirely, been substantiated.

So far as these developmental and personality traits are concerned, chronic alcoholics are not a definitely homogeneous group; a comparison of the differences of the alcoholics' self-estimates indicates that no one particular factor is characteristic of the entire group.

The following developmental factors and personality traits (either actual or so considered by the subject) differentiate the chronic alcoholics as a group from the controls.

a. A domineering but idealized mother and a stern, autocratic father whom the patient feared somewhat as a child.
b. A marked degree of strict, unquestioning obedience demanded in family life, with little latitude and freedom allowed.
c. A feeling of insecurity as evidenced by an insistent feeling of need for religious security and a strong feeling of sin and guilt.
d. Marked interest in the opposite sex, with many love affairs but poor marital adjustment.
e. Lack of self-consciousness, together with a marked ability to get along with and be socialy acceptable to others.
f. A keyed-up emotional level, resulting in work done under high nervous tension.
g. A definitely expressed and disproportionately greater love for the maternal parent than for the paternal.

Comparison of alcoholic with control answers would indicate that the self-estimates of chronic alcoholics appear to be influenced in general by:

a. A hyper-critical attitude toward others, together with a lack of insight into self-motivations.
b. An apparent "halo" effect.
c. A tendency to conventional sentimentality.
d. A defensive feeling of self-pity which shields the ego in adverse situations.

The experimental group as a whole show lower adjustment scores than the control group. Those divisions under which the alcoholics classify as poorly adjusted are concerned primarily with inner adjustments rather than adjustments between the individual and his environment, and with present adjustment rather than with childhood factors contributing to adjustment. This fact would seem to indicate that the alcoholic's lack of adjustment arises from within himself rather than from outside factors.

There is a large amount of overlapping between individual adjustment scores for the two groups. Maladjustment is found in the control group as well as in the experimental group, while, on the other hand, relatively excellent adjustment (self-judged, of course) is also found in both groups.

BIBLIOGRAPHY

1. BECKMAN, R. O. A new scale for gauging occupational rank. *Person. J.*, 1934, **13**, 225–233.
2. BENEDEK, T. Dominant ideas and their relation to morbid cravings. *Int. J. Psychoanal.*, 1936, **17**, 40–56.
3. CHAMBERS, F. T. Psychologic approach in certain cases of chronic alcoholism. *Ment. Hyg.*, 1937, **21**, 67–78.
4. CHASSELL, J. O. *The experience variables.* Rochester, N. Y.: Univ. of Rochester Medical School, 1928.
5. CURRAN, F. T. Personality studies in chronic alcoholic women. *J. nerv. ment. Dis.*, 1937, **86**, 645–667.
6. DAMRAU, F. Treatment of chronic alcoholism. *Med. Rec.*, June 15, 1938, 147.
7. FLEMING, R. Chronic alcoholism. *New Eng. J. Med.*, 1937, **217**, 779–783.
8. KNIGHT, R. P. Dynamics and treatment of chronic alcoholism. *Bull. Menninger Clin.*, 1937, **1**, 233–250.
9. MILES, W. R. Psychological factors in alcoholism. *Ment. Hyg.*, 1937, **21**, 529–548.
10. SHAW, C. R. *The Jack Roller.* Chicago: Univ. Chicago Press, 1930.
11. STRECKER, E. A. Psychology of alcoholism. *J. nerv. ment. Dis.*, 1937, **86**, 191–205.
12. STRECKER, E. A., AND CHAMBERS, F. T. *Alcohol—one man's meat.* New York: Macmillan, 1938.
13. TILLOTSON, K. J., AND FLEMING, R. Personality and sociological factors in prognosis and treatment of chronic alcoholism. *New Eng. J. Med.*, 1937, **217**, 611–615.
14. WALL, JAMES H. Study of alcoholism in men. *Amer. J. Psychiat.*, 1936, **92**, 1389–1401.
15. WALL, JAMES H. Study of alcoholism in women. *Amer. J. Psychiat.*, 1937, **93**, 943–955.

SHORT ARTICLES AND NOTES

THE ABILITY TO JUDGE CRIMES FROM PHOTOGRAPHS OF CRIMINALS: A CONTRIBUTION TO TECHNIQUE*

BY G. R. THORNTON
University of Nebraska

In the past it has been emphasized that external signs are unreliable as bases for judgments of personality. Experiments have demonstrated, for example, that shape of the nose, height of the forehead, distance between the eyes, size of the ears, etc., are not valid criteria of personality traits,[1] and that even men trained in personnel work cannot judge vocational aptitude from cues gained from studying photographs.[2] Attention has been turned principally to the practical problems of disproving the claims of charlatans and of demonstrating the unreliability of judgments based upon photographs. Little attention has been given to the equally significant, though less immediately practical, problem of discovering what, if anything, concerning personality characteristics can be discerned from these external criteria. Relatively crude techniques are adequate to establish the fact that judgments based upon external signs are too inaccurate to be of practical value. But to determine whether or not external signs yield any valid cues at all, requires more exact methods of measurement. The present study presents a simple method which yields measurements sufficiently exact to meet this need.

Problem

The University of Nebraska has for years included in the laboratory course in psychology an experimental study intended to demonstrate to the student the unreliability of judgments of character when they are based upon static external signs.[3] In one part of the study the students, shown photographs of criminals, were asked to judge what crime each man had committed. As the results from this experiment came in from group after group, it was noted that the average scores were consistently slightly better than might be expected by chance. The question thus became of interest: Do students have some ability to judge crimes by studying photographs?

Controls Exercised

In order to obtain an unambiguous answer to this question, it is necessary to rule out two possible sources of error. (1) Social stereotypes may influence the selection of the photographs used. (2) Factors inherent in the method of record-

* A preliminary report of this study was made before the Midwestern Psychological Association in April, 1938.

[1] G. U. Cleeton and F. B. Knight. Validity of character judgments based on external criteria. *J. appl. Psychol.*, 1924, **8**, 215–231.

[2] M. S. Viteles and K. R. Smith. The prediction of vocational aptitude and success from photographs. *J. exper. Psychol.*, 1932, **15**, 615–629.

[3] J. P. Guilford. *Laboratory studies in psychology*. New York: Henry Holt & Co., 1934, pp. 11 ff. The procedure used in this study is an adaptation of the procedure outlined by Guilford.

ing judgments may operate to cause some correct responses. These two sources may account for correct responses which nevertheless do not indicate any real ability to judge crimes from photographs.

Consider the first source of error. Suppose that in choosing the photographs the experimenter should unwittingly select cases which fit into certain stereotyped conceptions of various types of criminals. Similar stereotyped conceptions might cause the judges to name the correct crime. The results would indicate, then, not that the judges had ability to judge crimes from photographs, but simply that there was agreement between the stereotyped conceptions of experimenter and judges.

To avoid this source of ambiguity, the criminals in this experiment were selected on the basis of case records only. The authorities of the Nebraska State Penitentiary kindly permitted the experimenter to take from their files a number of records. From these, twenty were chosen, with regard for the following points: (1) each man had been convicted of only one crime, and so far as was indicated he had had no previous prison record; (2) each was white and a male; (3) the group of twenty represented a variety of crimes. The important point is that the cases chosen were selected on bases other than photographs; indeed, the photographs were not seen until after the selection had been made. The photographs used were those taken by the prison officials at the time when the men were incarcerated. From these, lantern slides were made which showed only the head and face of the individual.

These slides were presented to fourteen groups of students in a laboratory course in psychology. Each student made his judgment independently by underlining one of four possible crimes for each photograph shown. The record sheet which each judge used was developed in order to avoid the second of the two sources of error mentioned above: namely, that factors inherent in the method of recording judgments may cause a certain number of correct responses. In an experiment of this sort, the problem is to determine how much the cues from the photographs aid the judges in making correct responses. In order to determine this point, it is necessary to have a rather accurate determination of the number of correct responses occurring from other causes.

The assumption that these other causes will operate in a chance fashion is not an adequate solution of the problem. In a previous similar experiment at the University of Nebraska, in which the judges chose one of four answers, it was found that certain incorrect items were chosen much more frequently than other incorrect items. It was found, also, that the judges had a tendency to select the second item in the list more often than the other items. In other words, in this experiment other causes were not operating in a chance fashion.[4]

Nor is the problem solved by having the judges simply name the crimes. Under this procedure each judge chooses one name from an unknown number

[4] A simple control experiment yielded similar results. Sixty-three students checked one of four crimes in each of twenty lists, supposedly to test the possibility of mental telepathy. In the series of trials the following crimes each occurred three times: rape, libel, arson, embezzling, and dope peddling. Rape, libel, and arson were chosen a total of only 26, 30, and 30 times, respectively. Embezzling was chosen 53 times and dope peddling 51 times. The crime of fraud, which occurred only twice in the series, was chosen 46 times.

There was evident in this experiment also the tendency to choose the second item in the list of four. An interesting problem is whether this tendency to choose the second item is a function of particular items which occur second or a general tendency to choose the second in a list of four. Further experimentation with the position of the items systematically varied might easily determine which is the case.

of such names. Determining tendencies still operate to favor the recall and selection of certain responses. Under such circumstances the experimenter is even less able to estimate how frequently correct responses may occur from causes other than the one he wishes to measure.

The record sheet and its method of use in this experiment, it is believed, provide an adequate solution to the problem. The technique used is an adaptation of the familiar control experiment. The items appearing on the sheet were selected in such a way that for every correct item there appeared a control item which was in all respects equally likely to be chosen—except in so far as the judgment was determined by cues arising from the photographs. Thus any statistically reliable difference between the total score for correct items and the total score for control items could be attributed only to the influence of the photographs.

The method employed may be made clear by an analysis of the items. The following is a sample of the lists of items that appeared on the record sheet:

1. [arson], swindling, manslaughter, kidnapping
2. embezzlement, hold-up, [assault], trespassing

. . . . . .

10. dope peddling, fraud, incest, [robbery]
11. arson, swindling, [manslaughter], kidnapping
12. [embezzlement], hold-up, assault, trespassing

. . . . . .

20. dope peddling, fraud, [incest], robbery

The items that were the correct or control items in the original order have been underlined or boxed, respectively. The last ten lists on the sheet, those used in recording judgments in trials 11 to 20, were an *exact repetition* of the first ten lists. That is, the first list was the same as the eleventh; the second list was the same as the twelfth, etc. In each of the last ten lists, however, the control item was the one which previously had been the correct item, and the correct item was the one that previously had been the control. In list 10, for example, the correct item was *incest*, the control item *robbery*. In list 20, though the items were the same, the correct item was *robbery*, and the control item *incest*.

It is apparent, then, that in trials 10 and 20 the judge was presented with exactly the same choice, with the single exception that in the two trials different photographs were shown. Hence, if the results from a large group of judges should indicate, with statistical reliability, that *robbery* was chosen more often in trial 20 than in trial 10, and that *incest* was chosen more often in trial 10 than in trial 20, the only reasonable conclusion would be that something observed in the photographs was responsible for the more frequent choice of the correct items. Since all trials in the experiment were paired in the same manner, trial 1 with trial 11, trial 2 with trial 12, etc., it was possible to find a total score for correct items and compare it directly with the total score for control items.

One additional control needed to be exercised. There was a possibility that the order of the trials might tend to favor or oppose the selection of correct

items. This possibility was tested by changing the order of presentation of the photographs, so that photographs 11 to 20 became numbers 1 to 10. Thus any tendency to favor or oppose the selection of correct items that was present in the original order became a tendency to favor or oppose the selection of control items in the changed order—since by changing the order of photographs, those items which had previously been correct now became control items. The photographs were presented in the original order to the first 65 judges and in the changed order to the last 110 judges. As will be seen, the results are in the same direction for both orders of presentation; hence, whatever effect the order had, it does not change the general interpretation of the results.

In setting up the record sheet various other controls were exercised. The correct items (and likewise the control items) appeared five times in each column, i.e., first in the list five times, second five times, etc. In six cases in which the crime committed was the same for two criminals, the order of presentation of the item as a correct or a control item was the order A B B A. The items were located on the sheet so that the repetitions of the same item were about equally distant. Finally, the four crimes in each list were quite different from each other, in order to make the choice among them simpler.

TABLE 1

NUMBER OF SUBJECTS MAKING VARIOUS SCORES FOR CORRECT (C) AND CONTROL (X) ITEMS

SCORE	GROUP A		GROUP B		TOTAL	
	C	X	C	X	C	X
13			1	1	1	1
12			2		2	
11	1		1		2	
10	1		6	1	7	1
9	8		8		16	
8	10	4	19	6	29	10
7	5	6	21	10	26	16
6	17	8	23	16	40	24
5	9	14	10	21	19	35
4	9	14	12	23	21	37
3	2	10	7	23	9	33
2	3	6		4	3	10
1		3		4		7
0				1		1
N	65	65	110	110	175	175

RESULTS

The results here reported were obtained from 175 students in fourteen classes. The judges are divided into two groups, A and B, according to the order of presentation of the photographs, as explained above; the results are presented for each group separately and then for the two combined.

Table 1 presents the distribution of scores on correct items and control items. It will be noted that the trend of the scores in both the A and the B groups is higher for the correct than for the control items. Table 2 gives the mean

scores, standard deviations, and sigmas of the means for correct and control items. The means for correct items are in every instance larger than the means for control items. The magnitude of the differences, the sigmas of the differences, and their critical ratios are given in Table 3. Although the differences are not great—the greatest possible difference, if all correct items and no control items had been chosen, would be 20—these differences are highly reliable.

TABLE 2

MEANS, STANDARD DEVIATIONS, AND SIGMAS OF THE MEAN FOR SCORES ON CORRECT AND CONTROL ITEMS

GROUP	N	CORRECT ITEMS			CONTROL ITEMS		
		MEAN	S.D.	σ_M	MEAN	S.D.	σ_M
A	65	6.20	2.02	.253	4.51	1.79	.224
B	110	6.69	2.09	.199	4.68	1.97	.188
Total	175	6.51	2.08	.158	4.62	1.91	.144

TABLE 3

DIFFERENCES BETWEEN MEAN CORRECT SCORE (M_c) AND MEAN CONTROL SCORE (M_x), SIGMAS OF DIFFERENCES AND CRITICAL RATIOS

GROUP	$M_c - M_x$	σ_{diff}	C.R.
A	1.69	.338	5.00
B	2.01	.274	7.34
Total	1.89	.214	8.83

The results were analyzed also to see how many individual judges chose more correct than control responses. In both groups a large majority chose more correct responses. Of the total 175 judges, 124 chose more correct items, and only 40 chose more control items; 11 judges chose an equal number.

To summarize the results, we may say that, whatever method of analysis is used, the scores indicate a difference in favor of the correct items. This fact means that under the given experimental conditions most students show some ability to gain from photographs cues which aid them slightly in judging what crime a man committed. The ability, however, as found in this experiment is so slight that for *practical* purposes it is negligible.

DISCUSSION

The present study has two limitations which should be noted. (1) Only twenty photographs were used. The results obtained, therefore, might be due to chance agreements between certain of the photographs and the judges' stereotyped conceptions of certain types of criminals. Additional experiments with other photographs, chosen in the same manner as in this study, would overcome this limitation. (2) It is possible that some of the students had previous acquaintanceship with some of the criminals, since both the criminals

and most of the students were residents of Nebraska. That this limitation is not important seems probable because of two considerations. First, the penitentiary authorities permitted the experimenter to use only cases which had not had much publicity. Second, the student judges were asked to report the fact if they recognized any of the criminals; only one judge reported a recognition—and that proved to be a false recognition. A more adequate check upon this second limitation could be made by having persons in a distant locality serve as judges in the experiment.

If the results of this experiment are accepted as indicating that a person may get some cues concerning what crime a man committed from a photograph of that man, it becomes of interest to discover what are the cues which operate. The present study gives very little evidence on this problem; only a few pertinent facts may be noted. (1) It is possible that the age of the criminal may be an important factor. The criminals in this study ranged from 20 to 62 years in age; but it was not possible to pair them so as to determine the influence age may have had as a cue. (2) It is highly improbable that glasses or any ornaments or clothing acted as cues favoring the correct responses. The photographs used showed no clothing or ornaments; and only one man wore glasses. The latter was convicted of murder, but he was judged to be a murderer much less frequently than was the man who served as a control. (3) In some cases, cues led to erroneous judgments. In four cases (out of twenty) the control criminal, who had committed some other crime, was judged guilty of a given crime more often than was the man who had committed the crime. In two cases 115 and 123 judges agreed in judging a man guilty of some crime other than the one he had actually committed. That something in the photographs influenced these erroneous judgments is indicated by the fact that only 11 and 56 judges, respectively, chose these same wrong crimes for different photographs.

The present results suggest additional problems that might be investigated. (1) Are there individual differences in the ability to judge crimes from photographs? The range of scores in this study was rather large. Does this range represent chance variations only, or do some individuals have more ability than others in this task? (2) Can the ability to judge personality characteristics from photographs be increased by training? (3) Would the accuracy of judgments be increased notably by the use of moving pictures, or living subjects, as stimuli? In the study of these and perhaps other problems, the control techniques developed in this study might be used with advantage.

SUMMARY

The problem attacked in the present investigation was: Do students have any ability to judge crimes by studying photographs of criminals? Two important controls were exercised: (1) the twenty criminals were selected without any knowledge of their appearance; (2) a method of recording judgments was developed which made it possible to determine with some exactness how many of the correct responses should be attributed to cues arising from the photographs. It is this second control which is considered the important contribution of the study. The results from 175 students, who acted as judges, indicate that the average student can gain some cues from photographs which aid him slightly in making correct judgments. The number of correct judgments attributable to cues arising from the photographs is of negligible *practical* importance; but that the ability to gain such cues is real is attested by a high degree of statistical reliability.

THE JUDGMENT OF INTELLIGENCE FROM PHOTOGRAPHS *

BY STUART W. COOK
University of Minnesota

It is a generally accepted conclusion that an individual can estimate intelligence from photographs with no greater accuracy than may be expected on the basis of chance alone. Concerning the accuracy of the pooled estimate of a group of individuals opinion is divided; certain investigators believe the pooled estimate to be somewhat more accurate than that made by the individual, while others believe it to be equally inaccurate.

When the literature is examined, little justification is found for any definite conclusion on the problem without numerous qualifications. Although it is true that in the majority of studies the average (mean or median) of the individual coefficients of correlation between estimated and tested intelligence is less than .20, such is by no means always the case. In two of the most extensive studies this average is .425 (4) and .70 (12), coefficients representing agreement distinctly not of a random nature. Moreover, in all the studies using correlational techniques (with one exception), the range of individual coefficients, where reported, extends upward to the level of .80 and .90. This fact, along with the consequent implication that certain individuals *can* judge intelligence from photographs with a considerable degree of accuracy, has been overlooked in favor of the conclusion that all persons are inept at this practice because their average correlation (in most studies) is low.

With regard to the inaccuracy of pooled judgments, the evidence for the accepted generalization is even less adequate. While three studies (1; 6; 10) report a group coefficient below .30, three others (2; 4; 9) found coefficients above .40. One of these coefficients had a magnitude of .61 (2); another, one of .70 (4). It is true of all studies (2; 4; 9; 10) that comparisons between the average of the individual correlations and the group correlation are in favor of the latter.

Two factors of note are common to the various studies. First, with but two exceptions (1; 11), they made use of a small number of photographs and a large number of judges. Second, they did not use judges whose occupations gave them any claim to experience in the art of sizing up the mental characteristics of individuals on relatively short notice.

It is probable that the first of these factors accounts for both (1) the wide range of correlations within a given study and (2) the numerous discrepancies between the results of different studies. With regard to the wide range within a study, it is a familiar statistical fact that coefficients obtained from the repeated correlation of small numbers of scores from two variables will vary within wide limits about their true value. Thus, even if the true correlation between two such variables were zero, it would still be possible, with a distribution of only 10 or 12 cases, to obtain an occasional coefficient of the magnitude of $\pm .90$.

* For the suggestion and direction of this study as well as for counsel and assistance on many problems, I am indebted to the late Dr. C. R. Garvey.

just such a picture is found in many reports, i.e., a low mean or median coefficient with widely scattered extremes. This "curve of error," of course, indicates only a possible interpretation; what is really the case with these very high and very low coefficients must remain in doubt.

With regard to the discrepancies between studies, there are several possible explanations, all of which are directly related to the use of a small number of photographs. To illustrate, consider the unlikely possibility that for a certain small proportion of individuals there *is* a positive relationship between type of physiognomy and intelligence. It would follow that the obtained coefficient of correlation between these two variables would be a function of the number of such individuals included in the sample studied. Similarly, if certain "secondary cues" such as institutional hair cuts, slovenly dress, etc. were positively related to intelligence, the coefficient of correlation would be raised in proportion to the number of such cues not eliminated from the photographs. When small numbers of cases are used, influences such as these, if operative, are relatively important to the result. If large numbers are used, on the other hand, the true relationship for the total population is likely to mask such factors. It is possible, therefore, that, because the studies under consideration do have a small case frequency, discrepancies between them may be explained in terms of the nature of the sample of photographs.

The second thing noted about these experiments was that none of them made use of what might be called "experienced" judges. If conclusions about the problem are expected to appeal to those who are practiced in predicting ability from physiognomy, they must be drawn from results with individuals of at least an equal claim to proficiency.

PROCEDURE

As a consequence of these considerations, the present experiment made use of a total of 150 photographs and enlisted the cooperation of 10 "expert" judges. The subjects of the photographs included all the first-year male college students of the University of Richmond on whom data were available. This was, therefore, essentially a freshman group, including only a few transfer students. The photographs and the intelligence tests were both taken at the beginning of the year, thus insuring the presence in the distribution of a number of subjects, representing low extremes of intelligence, who probably would not have been in school a few months later.

The intelligence of the subjects was determined by the Thurstone Intelligence Test IV (No. 34063 Thurstone). The test was administered simultaneously to a large group of incoming new students by a professor of psychology who had had extensive experience in presenting it. The raw test scores are used in the correlations of this study.[1]

The photographs were of what is usually referred to as "passport" size (1.25" x 1.75"). They were uniform in every respect except dress. That this uncontrolled factor had no measurable influence on the judgments will be shown later under "Results." The part of the body extending from halfway between the shoulder and elbow upward was shown. All subjects sat in the same position, at the same distance from the camera, facing it squarely. No influence was exerted to have the subject record any given facial expression. Upon the back of the photograph there was placed an identification number chosen at random. The judges were informed of this random order of numbering.

[1] The raw scores range from 35 to 150.

Personnel managers and social workers were chosen as representing the occupations where the necessity for sizing up individuals in a brief period is most marked. Employers of personnel in an insurance company, a tobacco firm, an iron works firm, and the two largest department stores in the city, consented to serve as representatives of the personnel managers. Of this group one was a woman, while four were men. Likewise the aid of five well-trained, competent social workers was obtained. Three of this group were women and two were men. This gave a total group of judges, summed according to sex difference, of four women and six men.

Full typed instructions (a copy of which is given under "Instructions to the Judge") were presented to the judges with the packet of photographs. In order to make the instructions as uniform as possible from one judge to the next, no more was told him verbally than was absolutely necessary to make him understand the nature of what he was being asked to do. Each judge thus received, from identical typed sheets, the main portion of his information concerning the details of the technique he was to follow.

It will be noted from the "Instructions to the Judge" that care was taken not to restrict or limit the judgment on any one of the group of photographs. The judge was specifically freed of any stated obligations to place a certain number of photographs at a certain place on the scale. Each photograph was thus given a free range irrespective of the judgment on the photographs which had preceded it.

INSTRUCTIONS TO THE JUDGE

There are in the accompanying packet eight white cards, labelled from Group 1 to Group 8, eight rubber bands, a paper chart, and one hundred and fifty photographs, each bearing on the back a different number. These numbers are solely for the purpose of identification.

The young men who were the subjects of these photographs have been divided, according to their relative intelligence, into eight groups. Those with the lowest intelligence were placed in Group 1; those with the highest, in Group 8. The following diagram will illustrate:

GROUP 1	GROUP 2	GROUP 3	GROUP 4
Lowest intelligence	Just above lowest intelligence	Definitely below average intelligence	Just below average intelligence
GROUP 5	**GROUP 6**	**GROUP 7**	**GROUP 8**
Just above average intelligence	Definitely above average intelligence	Just below highest intelligence	Highest intelligence

The degree of difference in intelligence between the different groups is uniform. That is, Group 3 is as much less intelligent than Group 4 as Group 2 is than Group 3. Or, to put it the other way, Group 8 is as much more intelligent than Group 7 as Group 7 is than Group 6. As stated before, Group 1 represents those with the lowest intelligence, Group 8 those with the highest. Groups 4 and 5 would then contain, respectively, those just below the average intelligence and those just above. The remaining groups represent intermediate stages as indicated by the diagram.

The judge is requested to examine each photograph carefully, estimate the intelligence of the subject, and then place the photograph in the group corresponding to this estimate. For example, if the subject of photograph number 10 is estimated to be slightly below average intelligence, this photograph would be placed in Group 4. Likewise, if he were estimated to be just below the group with highest intelligence, the photograph would be placed in Group 7.

After judging of the photographs has been completed, those in Group 1 should be bound firmly with one of the enclosed rubber bands. The photographs placed in each of the other groups should be treated in a like manner. In order to avoid any possible confusion, the judge is also requested to note the number of the photograph in the proper place on the paper chart. Please separate the numbers—thus recorded—by short dashes.

The judge is reminded finally that the *intelligence* of the subjects is to be the *sole basis* for the separation of their photographs. The number of men placed in any one group according to their intelligence is not necessarily the same as that placed in any other group. The judge is therefore advised, in judging any one photograph, not to be influenced by the number of photographs which have already been placed in any particular group.

While no time limit is imposed on the judge, the return of the photographs is desired as soon as possible. The experimenter will either call for them or make other arrangements for their return.

RESULTS

1. As stated in the procedure, the photographs were placed by the judges in eight groups, the group with the lowest estimated intelligence being numbered "1" and that with the highest, "8." A corresponding division into eight categories was made on the basis of the raw intelligence test scores. By averaging the estimates of individual judges, a mean group placement for each photograph was obtained.

From these figures it was possible to check the effect of the uncontrolled variation in *type* of dress. 100 subjects wore coats and ties. The remaining 50 wore four combinations of sweaters, shirts and ties. When these two groups were compared, it was found that the arithmetic means of their group placements by the average judgment of the judges was almost identical—4.34 and 4.33, respectively. (The mean of the test intelligence of the two groups was also practically identical.) There is, therefore, no reason to believe that type of dress was an important determinant in the judges' decisions.

2. Product-moment correlation of the estimates of individual judges with test intelligence yielded coefficients which ranged from $-.06 \pm .055$ to $.20 \pm .053$, with a mean of .05. Nine of these ten coefficients ranged from $-.06 \pm .055$ to $.10 \pm .055$, none of the nine, of course, showing a statistically significant difference from zero. The tenth coefficient, $.20 \pm .053$, although it exceeds the .05 significance level and approaches the .01 level, like the others is of practically no predictive value. It may be concluded, therefore, that only one individual among these experienced judges could estimate intelligence from photographs with a reliably greater degree of accuracy than might be secured by a chance placement—and this one only slightly more accurately. This finding illustrates the point raised previously: that if individuals attempt to judge the intelligence of an *adequate sample* of photographs, results from *all* judges will be almost equally inaccurate.

3. The coefficient of correlation between the test intelligence and the average judgment, or mean group placement, of the judges on each photograph was $.07 \pm .055$. For those who consider school performance a more reliable criterion of intelligence than the intelligence test, a correlation of the average judgments with first-semester honor-point ratios was computed. The resulting coefficient was $.09 \pm .057$. Attention should be called to the fact that these coefficients do not support the frequent claim that the group will estimate intelligence more accurately than its individual members.

4. It should be pointed out that the judges were no more accurate in estimating the extremes of intelligence than they were in estimating average and near-average levels. As a matter of fact, the essentially random nature of the placements, as shown by the low correlations, resulted in larger rather than smaller errors on extreme cases. This fact may be seen directly by comparing the errors in placement on those photographs whose subjects fell into Group 1 and 8 on the basis of test intelligence with those that fell into the middle groups, 3 and 4.

The mean error of placement on photographs in the former categories was 3.05 groups, while the mean error for those in the latter was 1.50.

5. To what extent do judges agree on the judgment of intelligence from photographs? To answer this question a sample of 10 of the possible 45 inter-correlations among judges was taken. The sample was random except for the specification that each judge appear twice. The coefficients obtained range from $.15 \pm .053$ to $.45 \pm .044$ and have a mean of .28. While this finding indicates that the judges agree with each other to a noticeably greater degree than they do with the intelligence test, even this degree of agreement is to be considered very slight.

6. Though slight, the amount of agreement among the judges is sufficient to allow them to agree in placing some photographs relatively high and others relatively low. While there was no objective way of determining from these data what factors accounted for this tendency, a subjective analysis yielded tentative explanations. The ten photographs placed highest by the judges (mean group placement between 6.0 and 6.9) were carefully compared with the ten placed lowest (mean group placements between 1.7 and 2.8), and three very general points of difference noted.

The first of these may be best described as symmetry or regularity of features. In the high group no one feature was out of proportion to the rest of the face. With one exception, in which there was a narrowing toward the chin, the contour of the faces of the high group was of a normal, regular type, midway between oval and square. Only four of those in the low group approached this symmetry. In the low group, on the other hand, there were to be found individuals with one or more of the following characteristics: fat, oval face; narrow face; lopsided face; crossed eyes; large nose; large ears; thick lips; sunken eyes; overhanging eyebrows; and abnormally high eyebrows.

The second general factor noted was the position or appearance of the lips and the resulting expression given the face. Again with one exception—where a single individual wore a natural, attractive grin—the mouths of the high group were closed naturally and firmly, giving the photographs a quietly serious or mildly pleasant expression. At least eight of the low group, on the other hand, presented either bulging lips; slightly parted lips, giving a puzzled expression; or closed lips with the corners pulled away, giving a silly expression.

Finally, an even more general but very definite difference noted between the groups was neatness of appearance. While the presence of eight subjects with coats in both groups bears out the general conclusion stated previously that the type of dress had little to do with the judgments, it nevertheless appears that the neatness of the dress together with neatly arranged hair contributed a definite influence. The high group presented a uniformly neat appearance, the low group a uniformly untidy appearance.

SUMMARY

This paper reports an investigation of the success with which individuals of considerable experience in sizing up strangers in a brief period, can estimate intelligence from photographs. These individuals were asked to judge the intelligence of a large sample of 150 subjects on the hypothesis that the wide discrepancies in previous studies, both between individuals in the same study and between the averages of different studies, have been due to their use of a small sample. The following results were obtained:

1. *All* judges estimated intelligence with an approximately equal degree of

inaccuracy, the coefficients of correlation ranging from $-.06 \pm .055$ to $.20 \pm .053$ with a mean of .05.

2. The pooled judgment of the judges has an equally low correlation with intelligence, the coefficient being $.07 \pm .055$.

3. The judges were no more accurate in estimating extremes of intelligence than they were in estimating average or near-average levels.

4. Intercorrelations among the judges indicated a very slight degree of agreement, the coefficients ranging from $.15 \pm .053$ to $.45 \pm .044$ with a mean of .28.

5. Factors believed to be common to the ten photographs estimated to be of highest intelligence and missing from those estimated to be of lowest intelligence, are: (1) symmetry and regularity of features, (2) pleasantness of expression and (3) neatness of appearance.

BIBLIOGRAPHY

1. ANDERSON, L. D. Estimating intelligence by means of printed photographs. *J. appl. Psychol.*, 1921, **5**, 152–155.
2. COGAN, L. G. Judgments of character on the basis of photographs. As reported by Hollingworth, H. L., *Vocational psychology*. New York: D. Appleton, 1919. 48–53.
3. DASHIELL, J. F. *Fundamentals of objective psychology*. Boston: Houghton Mifflin, 1928, 306–307.
4. GASKILL, P. C., FENTON, N., AND PORTER, J. P. Judging the intelligence of boys from their photographs. *J. appl. Psychol.*, 1927, **11**, 394–403.
5. GURNEE, H. An analysis of the perception of intelligence in the face. *J. soc. Psychol.*, 1934, **5**, 82–90.
6. KNIGHT, H. C. A comparison of the reliability of group and individual judgments. Master's Essay in Columbia University Library, 1921. As reported by Murphy, G., and Murphy, L. B., *Experimental social psychology*. New York: Harpers, 1931, 504–505, 544–545.
7. LAIRD, D. A., AND REMMERS, H. A. A study of estimates of intelligence from photographs. *J. exp. Psychol.*, 1924, **7**, 429–446.
8. MARKEY, F. V. Variations in judgment. *J. appl. Psychol.*, 1934, **18**, 297–303.
9. MORIWAKI, E. Note on the comparative validities of judgments based on photographs and interviews. *J. appl. Psychol.*, 1929, **13**, 630–636.
10. PINTNER, R. Intelligence as estimated from photographs. *Psychol. Rev.*, 1918, **25**, 286–296.
11. POPE, D. V. The interpretation of the human face from photographs. *Bull. Randolph-Macon Woman's Coll.*, July–Sept., 1922, 3–17. As reported by Laird, D. A., and Remmers, H. A., A study of estimates of intelligence from photographs. *J. exp. Psychol.*, 1924, **7**, 429–446.
12. SCHMIDT, H. Äussere Erscheinung und Persönlichkeitsbeurteilung. (Judging personality from appearance), *Industr. Psychotech.*, 1934, **11**, 311–315. As reported in *Psychol. Abstr.*, 1935, 788.
13. UHRBROCK, R. S. Estimating intelligence from photographs. *Proc. Ninth Int. Cong. Psychol.*, 1929, 451–452.
14. WALLIN, J. E. W. *Clinical and abnormal psychology*. Boston: Houghton Mifflin, 1927, 556–559.

AN EXPERIMENTAL INVESTIGATION OF BIRKHOFF'S AESTHETIC MEASURE

BY DOUGLAS J. WILSON
University of Western Ontario

BEFORE presenting the methods and results of the present investigation it might be well to review the main features of Professor Birkhoff's theory of aesthetics as set forth very fully in his volume *Aesthetic Measure* (2). In contradistinction to other theories of art, such as "hedonistic, pedagogical, or mystical" (2, 191), the author describes his theory as "analytic," concerned entirely with the *formal* aspects of art which he explicitly states to be fundamental to aesthetics. The results of formal analysis are embodied in one basic formula which provides, for him, the index of aesthetic value. This formula is $M = \frac{O}{C}$, a mathematical way of stating that the aesthetic measure (M) of any stimulus is directly proportional to the elements of order (O) present in the constituent parts of the stimulus and inversely proportional to the complexity (C), which means merely the number of these parts. There are other ways of stating this conception that might seem more acceptable to the artist himself; the formula merely tries to state mathematically that the ultimate aesthetic measure of a painting, for example, is made up, on the objective side, of two major factors—complexity and orderliness. It further states that as the multiplicity of elements is increased, a skilful orderliness must be introduced by the artist to compensate for the complexity. Orderliness includes such factors as balance, various types of symmetry, equilibrium and so on.

Although the formula is intended to apply ultimately to all realms of aesthetic objects, in his book Birkhoff includes only examples from such stimuli as solid polygons, outlined polygons, various ornaments, tiles and vases, likewise music and poetry. In applying the basic formula it is necessary to alter the number and weighting of the constituents of O and C respectively, and Birkhoff admits that his suggestions are quite tentative. He has, in fact, already amended his original suggestion with respect to the evaluation of polygons (3, 353), though the change introduced does not affect any figure used in the present experiment.

Regarding this effort to arrive at a quantitative statement of the aesthetic value of an object, an effort which has been called by Pratt "heroic and ingenious" (6, 151), we may say that it offers certain challenges to psychological research.

First of all, it enjoys the advantage of being clear-cut and available for empirical testing. Not only has Birkhoff clearly defined and exemplified the meaning of each of the component factors in the basic formula, but he has included 90 polygonal forms with the aesthetic measure for each. Similarly, he has offered other visual and auditory material with his quantitative evaluations. Birkhoff's attempt has been described as a "step which all formalistic theories have needed to take, but apparently have not been bold enough to attempt, ever since the days of Plato" (6, 152). Accordingly, the theory escapes from the usual criticism of previous mathematical theories—namely, that of being vague and occult in statement and meaning. Professors Beebe-Center and Pratt express this point as follows: "Throughout the history of philosophy

various aestheticians, especially those who have inclined to formalism, have argued that the value of a work of art is primarily determined by the properties of the perceived object itself, but such argument has been largely a confession of faith rather than empirical demonstration. Birkhoff has had the courage to turn such faith into quantitative formulation and thus to bring it within the range of sensible argument" (1, 353).

Still more important is the fact that the logical or mathematical derivation of the formula is held to be secondary to psychological assumptions and justifications. A critical examination of these psychological foundations both thoroughgoing and, I believe, valid may be found in an article by R. C. Davis (4). Without reviewing Davis's criticism, I shall indicate what may be considered the major point of psychological controversy in Birkhoff's analysis, viz., that complexity is a summative process arrived at by mere enumeration of the parts; and further, that increasing complexity tends necessarily towards unfavorable aesthetic results.

Birkhoff's theory, then, is offered as one based on psychologically verifiable grounds, stimuli being provided by the author for such investigations. An attempt of this sort to offer a quantitative formulation that might conceivably make possible comparisons not only within one sphere of art but from one realm to another is, to say the least, interesting and daring.

Throughout his exposition, Birkhoff distinguishes between the "formal" associations and the "connotative" associations. The former refer to those associations which arise because of qualities in the stimulus that seem simple and unitary—such, for example, as symmetry. It is customary, I believe, to describe this type of response as "intuitive" in aesthetic literature, even though the formal bases of such immediate perceptions are capable of sometimes intricate analysis. The connotative associations, introducing into art the meaningful, symbolic factor, are so much dependent on private experience and past training as to render any kind of predictive formula highly precarious. In the present formula, connotative associations are supposed to enter only into the O (order) factor. It is Birkhoff's view that these associations can be largely eliminated, chiefly by regarding his polygons as actual tiles set up vertically. It would seem to the present investigator that Birkhoff thus invites a connotative association, namely, that of "tiles." In any case the suggestion proved inadequate in practice, as evidenced by the amended formula to which reference has already been made.

Previous investigations include those made by Professors Davis, Beebe-Center and Pratt. All of these have investigated the validity of the hypothesis as applied to various realms of art. Davis selected every tenth polygon offered by Birkhoff and had his 217 subjects evaluate the polygons by a ranking method. As a result he found "no support for the formula, either a priori or empirically" (4, 240). No reference need be made to Davis's results in other fields of art, though they were slightly more favorable to the theory, since the present investigation concerned itself entirely with polygons. Contrasted with Davis's conclusion is that offered by Beebe-Center and Pratt, based on a total of fifteen subjects. These investigators state that "clearly the results indicate that Professor Birkhoff's formula for polygons has a considerable degree of validity" (1, 341). The statistical evidence offered in this report failed to convince the present investigator that any such conclusion was either clear or possible. Later on, moreover, the authors "prefer to suspend judgment on the specific equation for polygons" (1, 342).

The research under discussion has differed somewhat from those referred to previously. In the first place, it is confined to the use of polygonal forms, both solid and in outline, because Birkhoff has stated that "polygonal forms constitute a class of aesthetic objects of the utmost simplicity, which have the further advantage of being relatively free from connotative elements of order" (2, 16). We may assume, therefore, that "independent aesthetic value" is at its maximum in primitive polygonal forms, that if the hypothesis could be sustained in any realm, it would be in that of simple polygons.

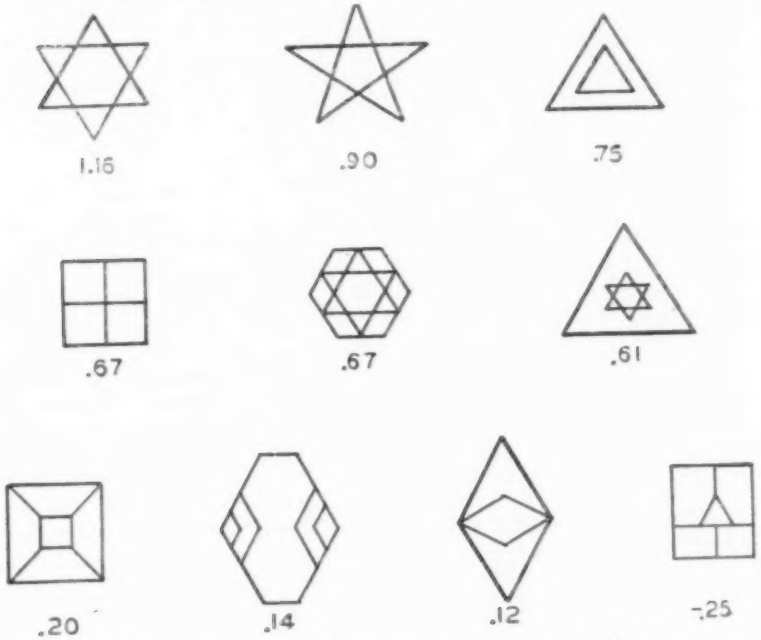


FIG. 1. OUTLINE POLYGONS USED BY AUTHOR IN PRESENT EXPERIMENT
Aesthetic Measure (M) is given for each.

A second difference in the present research from that of Davis was the method of selection from Birkhoff's figures. Previous routine experiments in our laboratory classes had followed Davis's plan of selecting every tenth figure. The only important difference in results we noted was that, whereas his subjects selected figure No. 50 as their first choice, ours consistently selected No. 40, a six-pointed star. The introspective accounts of these preliminary tests guided us in our selection of figures for the present investigation. In order to hold as constant as possible the "connotative associations" likely to be produced by heterogeneous polygons, we selected ten figures homogeneously composed, such as a symmetrical pentagon, hexagon, heptagon and so on. Not all the ten figures used were exactly of this symmetrical nature, but they could be described as being compact, regular, convex polygons without extraneous projections or re-entrant niches. All ten figures occurred in Birkhoff's plate, and his evaluations of them were known. A second series used consisted of figures in outline only, again chosen from Birkhoff by criteria that would tend to render all figures used as homogeneous in structure as possible within his series. (See Fig. 1.) By this more refined process of selection of stimuli it was hoped that we might overcome

Davis's failure to regulate in any way the connotative associations and offer the fairest possible opportunity for Birkhoff's hypothesis to receive empirical support.

A third difference from previous studies was the method of presentation. The stimuli, copied directly from Birkhoff's plate, in black rather than blue, were placed two on a card in a manner agreeable to the best requirements of the paired comparison method. Matters of position, temporal contiguity, repetitiveness and so on were carefully controlled. These figures were projected on a screen by means of a reflectoscope for approximately 10 seconds each. Accordingly, the position of the figures—not controlled by Davis, though demanded by Birkhoff—was in our investigation rendered vertical and constant.

Ninety-five undergraduate students in psychology, unfamiliar with the purposes of the experiment, acted as subjects. Each subject was provided with a blank containing at the top the following instructions:

Male..... Female..... Training in art, other than regular school curriculum: Yes..... No.....

This experiment consists of a series of judgments between various visual stimuli presented two at a time. Judge which one of each pair you believe to be preferable from the point of view of aesthetic beauty and record your preference opposite the appropriate number. Be sure to vote for one or the other; if there seems little choice vote for one and add a question mark in the column provided.

Try to assume a passive attitude and record your very first impression. Do not stop to analyse the stimuli. Avoid as much as possible associations or meanings that grow out of past experience. The question before you is, "Which of the two figures seems more aesthetically beautiful?"

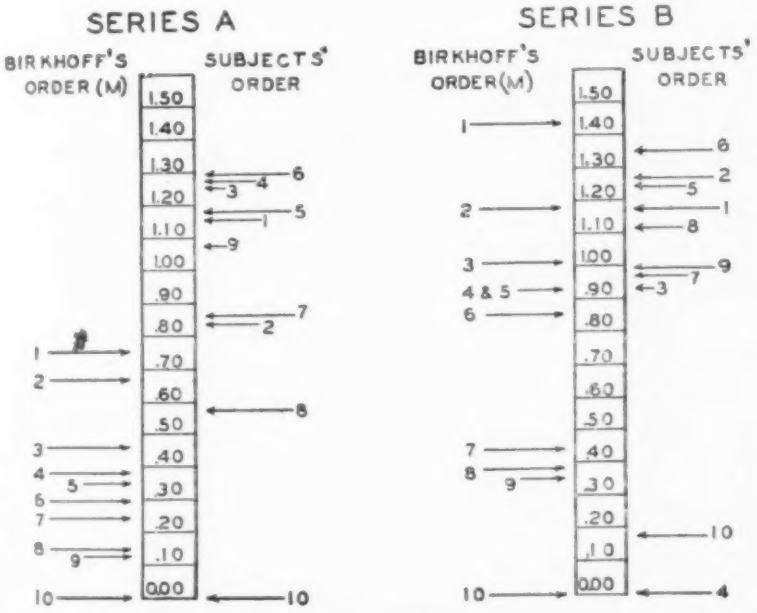


FIG. 2. GRAPHIC REPRESENTATION OF ORDER OF STIMULI ACCORDING TO BIRKHOFF'S MEASURE "M," AND THAT OF SCALE VALUES OBTAINED FROM CLASS OF 95 SUBJECTS BY METHOD OF PAIRED COMPARISONS

Note that Birkhoff's values have been corrected by eliminating all minus values and setting the lowest one at zero.

RESULTS

The results of the subjects' judgments were computed by the method proposed by Guilford (5, 236–238).[1] By this means scale values were obtained for both series A and B. A graphic representation of these results is given in Fig. 2. It should be understood that the actual position on the "M" scale for any stimulus ought not to be emphasized, as Birkhoff has repeatedly stated that his formula is not designed as a sensitive measure. The important point to note is the discrepancy in order for both series.

Another way of revealing the discrepancy between Birkhoff's *a priori* order and the order obtained in the present experiment is by correlating the two orders. When this process was carried out by the rank position method, the correlations for Series A and B were +.44 and +.38 respectively. In both instances the correlation between male and female was +.90.

These correlations are approximately what we have obtained previously using other sets of Birkhoff polygons.

A reference to the incidence and placement of interrogation marks, as required by the instructions, supports the above quantitative evidence. In general it is found that the greater the number of question marks for given comparisons, the closer together were the two figures according to scale value position, but not according to Birkhoff's order.

CONCLUSIONS

The results of this investigation do not support the hypothesis of Birkhoff that his *a priori* measure is a true measure of aesthetic value as far as polygons and geometrical figures are concerned. It is suggested furthermore that if the hypothesis fails here, it will fail *a fortiori* in more complex realms of art.

A more fruitful approach to the derivation of an aesthetic formula might be the attempt to weight those features that are found empirically to have most aesthetic value. For instance, triangles received consistently higher ratings than rectangular figures; yet triangles do not enjoy as high a rating by Birkhoff's hypothesis. Beebe-Center and Pratt have suggested an approach utilizing factorial analysis. Whether or not such methods ultimately will yield a formula for computing aesthetic value, for the present at least it seems as though Birkhoff's formula suffers from over-simplification and faulty psychological analysis.

BIBLIOGRAPHY

1. BEEBE-CENTER, J. G., AND PRATT, C. C. A test of Birkhoff's aesthetic measure. *J. Gen. Psychol.*, 1937, **17**, 339–354.
2. BIRKHOFF, G. D. *Aesthetic measure*. Cambridge, Mass.: Harvard University Press, 1933.
3. BIRKHOFF, G. D. The present status of esthetic measure. *Sci. Mon.*, 1938, **46**, 351–357.
4. DAVIS, R. C. An evaluation and test of Birkhoff's aesthetic measure formula. *J. Gen. Psychol.*, 1936, **15**, 231–240.
5. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
6. PRATT, C. C. Structural vs. expressive form in music. *J. Psychol.*, 1938, **5**, 149–156.

[1] $Mp = C + \frac{N}{2} \over nN$

Mp = Mean proportion
C = Total number of choices
N = Number of stimuli
n = Number of subjects

RELATIONSHIP BETWEEN INTEREST, ABILITY AND SELF-ESTIMATED ABILITY AMONG MAL-ADJUSTED BOYS

BY N. D. M. HIRSCH
Washington, D. C.

THE work of Thorndike is largely responsible for the suppositions that (a) interests are a dependable clue to ability and (b) that interests are permanent. He has concluded that it would be hard to find "any feature of a human being which was a much more permanent fact of his nature than his relative degrees of interest in different lines of thought and action." He has also affirmed that interest and ability are so closely linked together that "either may be used as a symptom of the other almost as well as for itself." Both of these theories are often accepted, although a number of studies have failed to confirm them.[1] The work of Professors H. Lehman and Paul Witty has done much in undermining the belief that the vocational interests of children are permanent; their studies suggest that "perhaps no phase of human nature is subject to such marked change as that reflected in the vocational interests and preferences of growing boys." They further suggest that since interests change, it is clear that "if ability be conceived as inborn and predetermined . . . the children's interests cannot be considered reliable indices of their abilities."[2]

If, however, even though the child's vocational interests are relatively transient there is an extremely high correlation between ability and interest, then nevertheless vocational bearings and guidance, to some extent, could still rely on the child's interests for aid in vocational orientation and training.

The subjects for the present investigation were boys referred to the Wayne County Clinic for Child Study, Detroit, Michigan, during the years 1930–1931 and during 1933. Most of these boys were juvenile delinquents, but the group also included some behavior problem and some dependent children.

Are children's interests and their abilities highly correlated? With this question in mind 600 maladjusted boys between the ages of 10 and 17 years, 6 months, inclusive, were given four tests, singularly unlike in form, content and method of working: The Binet, the Porteus Maze, the Healy Picture Completion II and the Ferguson Form Boards. At the completion of these four tests two questions were asked: (1) which test or game did you like best, which second and which third; and (2) on which test (or game) did you do best, on which second best and on which third best?

By this procedure (1) the correlation between ability and interest could be calculated, as well as (2) the association between ability and self-estimated ability and (3) the correlation between self-*estimated* ability and interest.

[1] Perhaps some of these failures in confirmation are due to the nature and level of the "interests" studied. It is a far step from interest in medicine or politics to interest in the four tests used in this research.

[2] E. K. Strong has shown that the interest patterns of individuals change greatly until between the ages of 25 and 30; from the latter ages on, they are relatively stable until 55 or 60 years of age. See his *Change of interests with age*, 1931.

The correlation between ability and interest, although positive, is judged to be rather low for the practical purpose of guidance, and taken in conjunction with the "transiency" of children's interests reduces the value of interests for vocational direction. There is better footing in respect to the second correlation. By asking a child how well, relatively, he did in this, that and the other game a fairly substantial agreement is discovered to exist between opinion and actual accomplishment. And there is a rather high correlation between the interests of these subjects and their self-estimated ability ($r = +.48 \pm .025$). Indeed interests seem a "symptom" of self-estimated ability rather than of real ability, but even here the estimates of it are very far from approaching a correlation of $+1$. Yet interests are much more an index of how a child rates or judges his abilities than an index of his abilities. And many juvenile interests, especially transient ones, are formed and developed by *wishes*, novelty, prestige and general halo. Wishes and immediate attractiveness, more than ability, generate a child's interest and in some measure control the child's appraisal of his ability.

TABLE 1

CORRELATIONS OF ABILITY, INTEREST AND SELF-ESTIMATED ABILITY

	ABILITY AND INTEREST	ABILITY AND SELF-ESTIMATED ABILITY	SELF-ESTIMATED ABILITY AND INTEREST
Degree of correlation	$r = +.20 \pm .029$	$r = +.39 \pm .027$	$r = +.48 \pm .025$

It may be further observed that there is a more positive association of ability with estimated ability than of ability with interest. In other words, the (maladjusted) juvenile's ability has even less association with his interests than it has with his self-judgment of his ability.

How does a child's degree of intelligence affect the three correlations, or how do differences in I.Q. affect the association (a) between ability and interest, (b) between ability and estimated ability, (c) between estimated ability and interest?

Scanning Table 2 from left to right, we note that the correlations between ability and interest are essentially unaffected by I.Q. status, and that a similar deduction may be drawn from the series of correlations between ability and estimated ability. The child's liking, preference or interest is not only very loosely associated with his ability but the degree of this association shows no increase with rise of I.Q. The practical implication here is that, on the basis of their juvenile interests, intelligent children can no more be vocationally guided than children of inferior mentality. The implication of the second row of correlations is that intelligence is apparently of but little aid in the self-evaluations of *children*. This fact is probably due to the introduction and interference of wishes and emotions. Studying the bottom row of figures, we find the correlation between estimated ability and interest indicates that I.Q. here is a more *positive* factor. The more intelligent child's interest and his self-estimated ability (belief in his ability) are more closely associated than is the case with the less intelligent child. But since the more intelligent child is but little better judge of his real ability than is the more mentally retarded child, vocational guidance by way of interest is an unreliable criterion. Special ability tests and tests of general intelligence instead of interests, transient or relatively permanent,

TABLE 2

RELATION OF BINET I.Q. RANGES TO THE THREE CORRELATIONS ON 600 CASES

	I.Q. RANGES							
	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	100 to 109	110 up
Number of cases	8	22	104	147	152	96	55	16
Ability-Interest	+.19±.029	+.23±.144	+.24±.064	+.26±.052	+.24±.052	+.21±.052	+.19±.062	+.27±.137
Ability-Estimated Ability	+.15	+.30±.138	+.33±.061	+.43±.046	+.36±.05	+.38±.057	+.42±.08	+.32
Estimated Ability-Interest	+.20	+.26±.144	+.34±.062	+.36±.05	+.36±.05	+.44±.055	+.51±.07	+.54

apparently are better guides in vocational selections and rejections for juveniles.

How does chronological age affect the three correlations? Do differences in chronological age radically alter the association (a) between ability and interest, (b) between ability and estimated ability, (c) between estimated ability and interest?

TABLE 3

RELATION OF CHRONOLOGICAL AGE TO THREE CORRELATIONS

	AGES			
	9-11 N = 71	12 & 13 N = 140	14-16 N = 349	17 to 17½ N = 40
Ability-Interest	+.26±.073	+.21±.053	+.19±.033	+.24±.102
Ability-Estimated Ability	+.31±.073	+.36±.05	+.37±.029	+.39±.089
Estimated Ability-Interest	+.58±.05	+.42±.05	+.44±.028	+.44±.085

Scanning Table 3 horizontally reveals that, although chronological age has little influence upon correlations between either ability and interest or ability and estimated ability, it affects the 9–11 year group markedly on the estimated ability-interest correlation; here $r = +.58$. This latter correlation suggests that youngsters below 12 are extremely apt to overestimate their ability in the field in which they are interested, or conversely, to be interested in those things in which they believe their abilities to be excellent. This suggestion is buttressed by the fact that the correlation between ability and estimated ability for this chronological group is only +.31. Emphasis too should be centered on the contrast between the ability-interest correlation of +.26 and the estimated ability-interest correlation of +.58 for this group.

TABLE 4

ASSOCIATION OF INTEREST, ABILITY AND ESTIMATED ABILITY IN A SECOND GROUP OF 300 MALADJUSTED BOYS

	ABILITY AND INTEREST	ABILITY AND SELF-ESTIMATED ABILITY	SELF-ESTIMATED ABILITY AND INTEREST
Degree of correlation	$r = +.20 \pm .037$	$r = +.37 \pm .032$	$r = +.50 \pm .029$

The same initial procedure involved with the foregoing work on 600 boys was followed two years later on 326 more cases, all maladjusted boys—an experiment undertaken as a check on the preceding study.

This supplementary work might be compared to Table 1 and the ensuing interpretation. The correlation again between ability and interest is perhaps of limited significance although positive; this relationship conjoined with the "transiency" of children's interests greatly reduces "interest" as a criterion for

juvenile vocational orientation. The other two correlations parallel those of Table 1, tending to confirm the hypothesis that interests are much more an index of how a child rates or judges his abilities than an index of his abilities.[3]

CONCLUSIONS

Not only are children's interests transient, but the correlation of such interests with ability is of a magnitude not generally accepted by psychologists as having much practical significance. A greater association is discovered between ability and self-estimated ability; and a fairly valuable correlation exists between self-estimated ability and the child's interest. A child's interests are, therefore, much more an index of his *wishful* estimate of his ability than of his real ability. This determination is true for various levels of intelligence as well as for varying chronological age groups.

[3] Since they are based entirely on work with maladjusted boys, these conclusions must be limited to them. An experiment of a similar nature on 500 to 1,000 unselected boys might have considerable value.

BODY FORM AND INTROVERSION-EXTROVERSION*

BY W. B. PILLSBURY
University of Michigan

Various results from experiments and more *ex cathedra* opinions have asserted a relation to exist between body form and mental characteristics of different kinds. There is, however, little agreement between the statements as to what the relation may be. In the hope that more light might be thrown upon the topic we correlated the Pignet Index of 274 male students with results of the Guilford form of the Introversion-extroversion Test.

The Pignet Index (height in centimeters, minus weight in kilograms, minus chest circumference in centimeters) was used as an index of body form because the data it demanded were available and because Kretschmer defined his three types in terms of it. It is at least objective and fairly constant for individuals of the same age and in a normal state of health. The measurements were taken from the records of the Department of Physical Education made when the students were admitted to college. The introversion-extroversion answers were weighted in accordance with Guilford's values.

The Pearson product moment correlation proved to be —0.019, which indicates a zero relationship between Pignet index and introversion-extroversion. At times, however, the coefficient of correlation is not always significant. Thus Cabot found only a slight correlation between body form and the results of the Bernreuter.[1] Nevertheless, a study of the individuals who were in the upper ten per cent of the several measures convinced him that extroversion was more prominent in the athletic type (athletosome) and introversion in the pyknic (pyknosome) and leptosome.

Similar methods applied to our data show little relationship. First we found that 4 individuals of the 10 per cent most extroverted (28) were among the 10 per cent most leptosome (with the highest Pignet Index). There were also 3 of the most extroverted among the 10 per cent most pyknic. Similarly, 3 in the most introverted 10 per cent were in the upper 10 per cent of the pyknic group, and one among the most introverted was in the highest 10 per cent in the leptosome group. This finding indicated that there was no significant relationship. No one of the 10 per cent most extroverted or most introverted was found in the 10 per cent most athletic.[2]

* The investigation was aided by funds from the N.Y.A. and by the University of Michigan Faculty Research Fund. Mr. Paul T. Forth, Mr. Bernard S. Rosenblat, and Miss Ann Sylvester should be mentioned for their aid.

[1] Cabot, P. S. The relationship between characteristics of personality and physique in adolescents. *Genet. Psychol. Monogr.*, 1938, **20**, 3–120.

[2] The athletic group were chosen by taking the 10 per cent closest to 17 in Pignet Index. Kretschmer defines the athletic group as those between 11 and 23 in their index. Thus 17 lies in the center of the group. Our 28 individuals ranged between 15.51 and 18.90.

A study of the average introvert-extrovert ratio for the extremes of body form showed what at first seemed a regularity. The highest 10 per cent in the Pignet Index had an average ratio of 1.26, the most pyknic (those with lowest ratio) had a ratio of 1.47, and the 10 per cent who were most athletic had a ratio of 1.83. This apparent regularity loses all significance, however, when it is seen that the standard deviations for the three groups were 4.44, 3.12, and 3.47, respectively, all many times as great as any difference obtained.

Moreover, the Pignet Index for the 10 per cent most introverted individuals was 2.44, and for the most extroverted was 2.15. These were opposed in direction to the previous apparent relationship. The most introverted were most leptosome, but the most leptosome were the least introverted.

No method of treating the results from these 274 students shows any relationship between body form as measured by the Pignet Index and introversion or extroversion as measured by the Guilford questionnaire.

DYSINGER'S CRITIQUE OF THE HUMM-WADSWORTH TEMPERAMENT SCALE [1]

BY DONCASTER G. HUMM
Los Angeles

In the original report of the standardization of the Humm-Wadsworth Temperament Scale the limitations of space required such close writing that it is evident certain explanations did not get full enough discussion.[2] For example, the fact that subjects were chosen by verified case study, including a careful review of their life history, for seven pairs of contrasting groups, seems to have escaped Dysinger.

Each pair of contrasting groups afforded a study of one of the seven components of temperament studied. The group of subjects evidencing a high degree of the "normal" component was contrasted with the group evidencing a low degree; the group evidencing a high degree of hysteroid, with the group evidencing a low degree; and so on for manic cycloid, depressive cycloid, autistic schizoid, paranoid schizoid, and epileptoid. The statistical data for these fourteen groups are given in Table VII of the report of standardization.

That all this was not understood by Dysinger is apparent from two of his statements: "Since Humm and Wadsworth do not give statistical data for the various components obtained from their group (*sic*) of subjects, it is impossible to make statistical comparison of the groups"; and, "It seems that Humm and Wadsworth divided the score on each component into two groups; one made up of those scores above the median or average and the other of those scores below the median or average."

Dysinger is of course quite right in concluding that his three groups of late adolescents may not be compared with these fourteen specially selected groups of adults used in the standardization. He could, however, have compared his groups with the group of unselected college students reported in Table Ia of the manual of directions which accompanies the scale.[3]

His conclusion that "the normal individual, then, is the one within the organization of whose personality these components are reasonably well balanced" naturally follows from his assumption that our subjects were chosen at random; but it is in error when one understands that the standardization procedure involved, not a study of population with a consequent report of trends, but the establishment of norms for certain groups of traits as observed in seven contrasting pairs of groups of individuals. It is to be noted that the report of standardization states, "Usually one component appears higher on the profile than any other," and that six cuts are included as illustrations.

In the report of standardization the validity of the mathematical procedures was tested, among other ways, by noting the correlation between groups of

[1] Dysinger, D. W. A critique of the Humm-Wadsworth Temperament Scale. This JOURNAL, 1939, **1**, 73–83.

[2] Humm, D. G., and Wadsworth, G. W., Jr. The Humm-Wadsworth Temperament Scale. *Amer. J. Psychiat.*, 1935, **92**, 163–200.

[3] Humm, D. G., and Wadsworth, G. W., Jr. *The Humm-Wadsworth Temperament Scale*. Los Angeles: D. G. Humm, 1934.

subjects in affirmative response to questions. The purpose of this correlation study was simply to determine whether the procedures had revealed that subjects of different temperamental tendencies were responding in distinctive manner. Dysinger compares those correlations with correlations between scores in the several components by individuals tested. His findings report the extent to which components appear together in given individuals. It does not seem as though there were a valid comparison between his intercorrelations and the correlations between groups of subjects as reported in the standardization report.

Dysinger does not report on the rectilinearity of his correlation fields. He does report very markedly skewed distributions when he reports by component points. There seems to be considerable reason to suspect that his correlation fields are curvilinear. If such is the case, he should either have corrected for curvilinearity or reported by correlation ratios. This consideration will be more particularly in point if he proceeds with factor analysis, as he intends, since involved manipulation may increase any inaccuracies that may be present.

During the standardization of the scale it was discovered that many individuals who had answered the questions with a high preponderence of No-responses were negativistic toward the scale to the point of setting up such defenses as lying, evasion, etc. It was also discovered that those who answered with very few No-responses were so suggestible as to imagine they had a given symptom when presented with a printed statement of it. Since individual case studies had already been made, it was simple to note that validity of response tended to diminish toward both extremes of the distribution of No-responses. A check was then made with the subjects to determine the attitude of each subject when answering the scale. This procedure resulted in three distributions on the same base line as reported in Table I of the report of standardization.[4] Accompanying this table is a recommendation for the acceptance or rejection of scales on the basis of No-responses. It would seem a mistake on the part of Dysinger to disregard these recommendations without establishing their lack of validity.

We are grateful to Dysinger for his retest reliability of each component and for his correlations between the components and Army Alpha scores.

Many of the points raised by Dysinger and briefly noted here are further discussed in a recent study in which I participated.[5]

[4] *Loc. cit.*

[5] Humm, D. G., Storment, R. C., and Iorns, M. E. Combination Scores for the Humm-Wadsworth Temperament Scale. *J. Psychol.*, 1939, 7, 227–253.

REVIEWS

EDITED BY EDNA HEIDBREDER

THE BEHAVIOR OF ORGANISMS. By B. F. Skinner. New York: Appleton-Century, 1938. Pp. ix+457.

To review Skinner's book one must first be certain of the author's purpose. True of any book, this requirement becomes a special problem with *The Behavior of Organisms,* inasmuch as Skinner himself both affirms and denies certain objectives. Has he attempted to give us a summary of experimental work in a specialized field, or has he attempted to present a psychological system? Early in the book Skinner points out that he is "interested, first, in setting up a system of behavior in terms of which the facts of a science may be stated and, second, in testing the system experimentally at some of its more important points" (p. 5). Then, after completing his argument, Skinner becomes unbecomingly cautious for a system-builder and requests us to remember that "the book represents nothing more than an experimental analysis of a representative sample of behavior. *Let him extrapolate who will"* (p. 442; italics my own). Despite this warning, we feel that Skinner has really attempted the construction of a generalized psychological system, and as such we shall attempt to review the book.

First, however, we wish to say a word with reference to the author's secondary purpose—the presentation of his experimental results. These data mark, we believe, a new "high" in experimental achievement. With the qualification that the statement refers to a sharply defined "representative sample of behavior," this reviewer can agree with the publisher's comment that the book "represents the most successful description of the behavior of the individual organism thus far achieved." The regularity and lawfulness of behavior which Skinner discovers should serve as an inspiration and guide to any experimenter. A detailed discussion of these data will not be attempted here, but we do wish to call specific attention of the reader to Skinner's work on the "reflex reserve," spontaneous recovery, and the hunger drive. The data accumulated there cannot be ignored in any future discussion of the learning process.

Of Skinner's system we must be more critical—critical not only of his systematic formulations, but also of his procedure in arriving at them. Practically all of his formulations are based upon data obtained in conditioning and experimentally extinguishing the "response of pressing down a lever" in the white rat. At almost no point does he seem to have been aware of the rather overwhelming fact that hundreds of other experiments have been done in the field of psychology. Such neglect of the work of other investigators might conceivably be excusable in a specialized monograph, but is hardly understandable in a book purporting to present even an *outline* for a generalized system of behavior. The result is that at very many points Skinner's conclusions and data stand in unexplained conflict with the conclusions and data obtained by other experimentalists.

We mention only a few instances. His espousal of the thesis that discrimination-learning is a continuous function due to a cumulative reinforcement and extinction of the positive and negative stimuli respectively (p. 169) neglects

entirely the results of a great deal of contrary evidence from other laboratories. Again, in discussing the stereotyping of responses and in formulating a "Law of Frequency" to describe this process (p. 309), Skinner comes to conclusions directly opposed to data accumulated by Muenzinger (using, what is more, an experimental technique very similar to Skinner's) and reported as long ago as 1928. And still again, in his discussion of the effects of reinforcement on the acquisition of a response, he neglects all the relevant work of Tolman and his students on the phenomenon of latent learning, where responses seem to be acquired *without* specific and differential reinforcements. In some instances this disregard for psychological data leads the author to very queer assertions. Thus in discussing the difference between extinction and forgetting he writes, "Forgetting is a loss of strength due, not to the active process of extinction, . . . but to the passive lapse of time" (p. 91). We know of no contemporary thinking on the problem of forgetting which would justify such a statement.

About the only "psychologist" whose work Skinner mentions (aside from his own co-workers) is the physiologist Pavlov! If Skinner were interested in setting up a System-of-the-Behavior-of-the-White-Rat-in-the-Skinner-Box, that would be one thing; but if, on the contrary, he is interested in setting up a system of *behavior* (and again, in self-protection, we must quote him: "The central problem of the book is the formulation of behavior, the treatment of which is intended to be exhaustive," p. 46), then his exclusion of psychological data and problems—e.g., perception, thinking, etc.—not stemming directly from his own work is a fatal error. The result is a "system," the laws and generalizations of which are extremely limited in usefulness to psychologists. In this connection it is very queer indeed to find Skinner dismissing Crozier's "system" because it lacks generality (p. 435)!

So much for Skinner's sins of omission. We might now examine some of the guiding principles of his system. First, he makes a convincing plea for a system of *behavior*. He wants psychology to study behavior *qua* behavior, without recourse either to the C.N.S. (Skinner refers to this as the "Conceptual Nervous System") or to mental entities. In so doing he is following in the tradition of most modern systematizers—Lewin, Tolman, Hull.

Second, Skinner intends to set up a purely *descriptive* science, one which is to avoid "explanations" and hypotheses. This is neither the time nor the place for an inevitably futile discussion of the assumed differences between "descriptive" and "explanatory" sciences. I wish merely to paraphrase another reviewer in this connection. Despite Skinner's intention to hew to the line of the strict positivist and operationist, he really does not (and I would say he *cannot*) get along without hypotheses. The "laws" which Skinner uses "direct the inquiry and hence are surrogates for hypotheses."[1]

Finally, our author proclaims his system to be a "molecular" and "analytical" one. "There is no reason," writes Skinner, "to appeal to any unique property of the whole sample as an 'act.' I make these statements as explicitly as possible in view of prevailing opinions to the contrary" (p. 55). Again, on page 29, he says:

> Interaction may be studied in a practical way by deliberately combining previously isolated units and observing their effect upon one another. In this way we obtain a number of laws which enable us to deal with those larger samples of behavior sometimes dubiously if not erroneously designated as "wholes." That great pseudo-problem—Is the whole greater than the sum of its parts?—takes in the present case this intelligible form: What happens when reflexes interact?

[1] E. R. Hilgard. *Psychol. Bull.*, 1939, **36**, 121–125.

Obviously Skinner is making a valiant attempt to differ with the Gestalt psychologists, but nowhere in his 450 pages does he ever call his self-selected adversary by name. There are, indeed, only two references in his bibliography to publications of Gestalt psychologists. And as to the "great pseudo-problem" which Skinner demolishes with his "intelligible" approach, we would ask, *Whose* pseudo-problem? Certainly the Gestalt psychologists have never formulated such a problem. Despite his attempt at a quarrel, however, we feel that actually Skinner does *not* differ from the Gestaltist—at least not in the way he thinks he does. We will assert that Skinner does not treat his data in a "molecular" fashion (see Skinner's assertion to the contrary on page 52 of his book). The analytical unit Skinner adopts is the reflex. But it is a very different reflex from that castigated by the Gestaltists. First, Skinner redefines the reflex so as to remove from it any neurological implications. A reflex, for him, is merely an observed correlation between stimulus and response. His definitions of these two fundamental terms are then carefully presented and should be placed on the "required reading" list of all "saliva-drunk" psychologists, as Tolman calls the conditioned-response experimentalists. A response Skinner defines in terms of a *class* of behavior, rather than in terms of the muscles involved. Actually, this is defining a response in terms of *what is accomplished*, in terms of the "meaning" of the muscle-movements. Thus:

> The response of pressing the lever is defined as any movement causing the lever to move downward. Some instances may be of unusual form. For example, the rat may stand on its hind legs, lose its balance, and fall against the lever (p. 70).

The stimulus is also defined in terms of its functional value for behavior. Skinner, however, has another criterion for a response. He wishes to narrow down a response as specifically as is possible. The range of behavior-acts which can be included within a definition of a given response is to be determined by the *lawfulness* of the relationship between stimulus and response; that is, in delimiting the response and the stimulus, one must leave the definitions of these terms just broad enough so that an orderly relationship can be discovered between them—and this relationship determines the existence of a psychological unit: the reflex.

Such a definition of a unit may have its merits, and certainly in Skinner's work it has shown itself to be extremely useful, but, disregarding the possibility of other and perhaps more generally useful units, we wish merely to emphasize the point that Skinner's reflex is *not* a "molecular" unit, as understood and criticized by the Gestalt psychologists. Gestalt psychologists never have objected to "psychological units" and to analysis. That is a misunderstanding all too common among American critics. What the Gestalt psychologists *have* criticized is the type of unit used and the degree of analysis indulged in, but certainly not the fact of analysis. They have maintained that each science must determine for itself what its fundamental unit is to be, and that the appropriate unit need not always be the smallest discernible "molecule."

Skinner himself follows this fundamental position of the Gestalt psychologist. Thus, for instance, in considering his "chain of reflexes" Skinner sees that "the behavior of chewing and swallowing may be analyzed into a rather intricate series of reflexes, *which are only of slight interest here*" (p. 52; italics my own). Why does he, champion of the "molecular analytical method" that he is, not analyze the rather "intricate series of reflexes involved?" The physiologist certainly does so. The answer is simple. Skinner's reflex, as observed above,

is a *molar* one; the physiologist's, a "molecular" one (relative to Skinner's). Again we find Skinner following in the best traditions of Tolman, Lewin, Koffka, *et al.;* for if he stops with his process of analysis at a point beyond which no *psychological* sense can result, he is not differing from the Gestaltist's concept of what is proper analysis. His attempt to differ, however, leads him at several points to self-contradictory analyses. Thus, on page 106, he concludes that a certain experiment "shows the autonomous status of each member of the chain very clearly. It would be difficult to reconcile data of this sort with a doctrine of the 'wholeness' of the 'act of pressing the lever to get food.'" Yet, on the preceding page he has explicitly made it clear that the very existence of any member of the "chain" depends upon the "ultimate" reinforcement, i.e., in this case the obtaining of food. Certainly a single member of a chain which is sensitive to the "ultimate" effect, an effect which occurs several reflexes removed from the one under consideration, cannot be considered as an "autonomous" one!

Finally, we must end this review with a word of general regret. Because of the obviously important data Skinner has accumulated, because of the clarity of some of his thinking and the sharpness of his criticism of the traditional sterile conditioned-response concepts, it is to be profoundly regretted that Skinner did not attempt to tie up his work and his thinking with the rest of psychology's experimental data and concepts. He should have given us a more "social" and a more generalized product. This reviewer believes that such a product would have been a really significant contribution to systematic psychology. As it is, we have a collection of data which have yet to be patterned (together with other and just as significant data) into a useful system describing the behavior of organisms.

I. KRECHEVSKY.

University of Colorado.

THE REDISCOVERY OF MAN. By Henry C. Link. New York: The Macmillan Company, 1938. Pp. ix+257.

Like many another recent work, this book is concerned with the relationship between personality organization and social organization. Viewed from one angle its thesis is disarming and familiar. Through his studied development of an intellectualized and scientific culture, man has produced a pseudo-civilization in the Frankenstein pattern which now threatens to enslave and destroy him. Faced with the necessity of dealing responsibly with this situation, he incurs disintegration and shrinkage of personality and loudly blames every cause but his own ineptitude. Now on the verge of losing his last hold on self-respect, he is in need of rediscovery. Fortunately, rescue is at hand in the form of psychology (the author's version) allied with social conservatism, physical culture, sportsmanship, and the old-time religion.

For all the blatancy and bluster of its presentation, there is much truth in the general conception. Those familiar with the recent writings of Jung will acknowledge Link's diagnostic interpretation to be largely justified, although they will doubtless deplore the proposed cure. Viewed from another angle, however, many of the truths stand revealed as half-truths. That this is extensively the case, and that the general reader is unlikely to be aware of it, are matters best left to the sociologist, the economist, and the educator, all of whom receive at Link's hands a not altogether unmerited chastising. Here, however, it is more pertinent to consider the psychological aspects of the thesis.

Psychology is defined as "the scientific study of habits and their formation." Consequently the extrication of modern man from his predicament depends on procuring the formation of those habits which best adjust him to life, chiefly in terms of his responsibility to himself and to others. Such an end may best be achieved by lopping off intellectual habits and adding more of the elemental habits which derive from the strenuous push of life. For determining the procedure, "the true issue is: *What habits and habit-forming activities should be compulsory?*" The answer is: competitive activities of all kinds, especially athletic. Psychology also includes emotions, but their laws are the law of habits.

As might be expected, the treatment of personality, like that of psychology in general, is a triumph for the doctrine of functional specificity. If the earlier proponents of this doctrine seemed radical, Link here represents a sort of Lenin for whom competitive determinism is implemented by the dictatorship of the extravert. Personality is "the extent to which the individual has developed habits and skills which interest and serve other people." The nature of the habits and skills is determined not by self-interest, however, but by some meta-social nexus of values known as "the rules of the game"; these undergo historical transmission by such varied media as the Ten Commandments, the life of Jesus (who was "the supreme personality"), and the playing-fields of Eton. The needful habits may, moreover, be classified as: habits of social initiative, habits of self-determination, and habits of economic self-determination. Any and all forms of introversion are to be avoided as opiates of the psyche.

The principal implications of this exposition seem to be that if there were only extraverts in the world, the introverts would be less of a problem, and that the best kind of personality to have (unless, like Lincoln, one is gifted with a "super-personality") is a *standardized* personality. Standardization is achieved by following the rules of the game, and if the concept (which the author, to be sure, does not call by this name) seems somewhat contradictory, the reader may be reminded that nowhere in Link's treatment of personality is any adequate provision made for such qualities as individuality, originality, imagination—nor for that matter, for much of anything recognizable as mental cultivation. If only the right habits are forced on the individual as by mass-production (somewhere in the book the family is described as "a little Fascism"), competition meanwhile being maintained for its own sake, the distribution of the P.Q. will be skewed to the side of bigger and better personalities for all. Such an enlarged personality represents man rediscovered.

It follows as a matter of course that the rediscovered man, who has found himself through learning how to give others what they say they want in a sportsmanlike manner, will be a rugged individualist (without individuality). He will prefer personal security to social security, a high personality quotient to a high intelligence quotient, old-time virtue to risky experiments, and—in a world in which "geology and evolution portray [man] as little more than a monkey" (page 9)—the old-time religion to newfangled progressive education. He will likewise prefer "gold standards of personality" (title of Chapter IX) and all other gold standards, for the rediscovered, competitive man appears to be the Commercial Man *par excellence*. And, exulting in his new-found freedom from bondage, he will ignore alleged defects of the *status quo* except for occasionally patronizing the shiftless poor who create the slums. In all respects but one he will behave like an old-fashioned utilitarian. The novel procedure lies in playing the game for its own sake rather than for himself.

It is a deep truth of life that salvation is secured by sacrificial unselfishness.

It seems equally true that the world can be no better integrated than the individuals in it. Dr. Link acknowledges these truths and makes them the foundation of his teaching. But his presentation and development, including as it does the recognition of religion as "competition of souls," is a travesty phrased in half-truths from which emerges no emancipated man but rather an amiable neurological jumping-jack completely conditioned to the patterns of extraverted behavior.

HOWARD DAVIS SPOERL.

American International College.

PSYCHOLOGY OF INDIVIDUAL DIFFERENCES. By A. R. Gilliland and E. L. Clark. New York: Prentice-Hall, 1939. Pp. xvi+535.

"The purpose of this book," announce the authors in their preface, "is to collect in one volume the more important facts and conclusions in the broad field of individual differences." This might sound as if a handbook were intended, but examination shows the volume to be definitely, frankly,—and on the whole successfully—a textbook for students just beginning to survey the field concerned. Properly enough, therefore, emphasis has been laid not upon inclusiveness, but rather upon selection of what the writers consider "important," upon organization, and upon a simple compact style.

The writers array their material under the usual, perhaps inevitable, headings: I. Problem; II. Causes; III. Measurement of Differences; IV. Differences in Physical Characteristics; V. Sex Differences; VI. Race Differences; VII. The Influence of Near Ancestry; VIII. Differences in Intelligence; IX. Types of Extreme Deviations; and X. Differences in Personality. Practical applications of the lore are discussed in: XI. Individual Differences in Learning and Teaching; XII. Applications of Individual Differences in Business and Industry; and in XIII. The Significance of Individual Differences. It is not necessary to summarize here the contents of the majority of these sections. In general, they cover the subject in a clear, factual, fairly orthodox way, at the level of the student who has been exposed to no more than an introductory course in general psychology. Since the field is one which is familiar to the readers of this review, it is sufficient and more profitable to draw attention here primarily to some of the book's more individual aspects.

One rather striking departure is the divorce of "Influence of Near Ancestry" from the discussion of heredity and environment as causes of individual differences—a discussion which it follows belatedly after an interval of four chapters. It appears that the authors believe studies of relationship and of family background to be in danger of too ready acceptance as evidence of hereditary factors; and they prefer, therefore, to hold them apart from the "nature-nurture controversy," and to allow them to follow "Race Differences," on which, undoubtedly, they have appropriate bearing. Many, however, even while sympathizing with the authors' apprehensions, are likely to find this procedure arbitrary. It is somehow hard to believe that the new arrangement will confuse the student any less than the traditional one.

The useful section on "Measurement of Differences" is concerned with mathematical concepts such as distribution curves, measures of central tendency, variability and correlation. It gives, in clear, non-technical language, an outline of essential points with which the student needs particularly to become acquainted. Since actual formulae are carefully avoided to make the presentation more palatable and easily digestible, the information will of course have to be

very heavily supplemented in order to become of very practical service, but it is perhaps something that an introduction, at least, has been effected.

Unlike the usual chapter which one has come to expect, that on "Individual Differences in Learning and Teaching" is concerned with tests in school subjects only as their results reflect the wide variations in ability which different pupils exhibit. Lock-step teaching methods are compared with more progressive ones, and various techniques—ranging from "flexible promotion," through the Winnetka and Dalton plans, to college honors courses—shown to be devices by which instruction can be adapted to meet individual needs. In view of the fact that a large number of those who will use this text are probably headed for the teaching profession, these practical considerations are not irrelevant and should be of appreciable value.

Perhaps the most notable omissions occur in connection with intelligence and intelligence testing. It is, to be sure, in line with the authors' policy of avoiding controversy that definitions and theories of intelligence are touched somewhat briefly and delicately, but it is curious that, whereas tests of personality are given intensive consideration, those of intelligence are awarded a scant four pages in the introductory chapter on "the problem," and are not reexamined. Possibly it is assumed that the student is pretty sure to meet these in another college course. Yet it would seem highly desirable that he should have had some of the actual tools of measurement explicitly described and explained, before he can be expected to understand or evaluate the differencs which the authors report in their comparisons of individuals, sexes, and races.

Great care has been taken to render the presentation consecutive and lucid. The points are made logically, one by one, each in its own sentence—the frequency of periods, especially in the early pages, serving doubtless as a reminder that the contents are to be studied rather than merely read. Another pedagogical device, of aid to the student and hence to the instructor, is the inclusion of questions on the text, following each chapter. The exposition is effectively illustrated with graphs, cuts, and tables. Footnote references at the bottom of the page, not grouped in a separate section, encourage the interested student to further exploration.

Throughout the first ten chapters, the authors adhere firmly to an impartial, non-controversial attitude wherever practicable. Dutifully they hold themselves down, for the most part, to a factual report of data, refrain from very basic interpretation of theory, and cautiously conclude the majority of sections with reminders that the problem is still open, that there is much to be said on two sides, and that the whole subject is rather involved. This is a fair approach, as well as the part of safety, and may presumably stimulate the student by suggesting that there is further work to be done. It must be admitted, nevertheless, that the treatment sometimes seems to lack the organization offered by theory-patterns, and the reader may wonder whether some of the conflicts could not be further interpreted, even when they cannot be finally decided.

On the other hand, one integrative point of view *is* found implicit throughout, and toward the end of the book is frankly emphasized. This stand, that a study of differences should lead to practical ends, involves the credo that systems and institutions which adjust to individual differences are, for that very reason, sounder and the more desirable. The authors are assured of wide support in this view when they discuss educational and industrial applications, but they find themselves in a less comfortable position when they apply the same rule to matters social and political. "When people come to realize this fact [that people do differ] more fully, and to organize social groups on the basis of

ability and interests, we will begin to have a better society." The same basis is urged for political structures, the authors sketching successive types of organization, each of which displays a higher level than the last in its appreciation of individual differences. Their list then stands: "Pure Democracies," "Communism," "Socialism," "Monarchies," "Republics," and "Fascism." Discovering, however, where such a train of thought seems to lead, the writers quickly explain that Fascism rests on differences bred of "force and coercion"; ultimately they give the seal of approval to Republics, which "accept the fact that people differ." Equality before the law and in voting is condoned in view of the fact that the election of men to office does create political differences, that "industry and ability may be rewarded by the highest office in the land." Wistfully the authors add, however, "it need only be mentioned that this is the ideal, which is never attained in practice."

EDITH B. MALLORY.

Wellesley College.

THE SOCIAL LIFE OF ANIMALS. By W. C. Allee. New York: W. W. Norton and Company, 1938. Pp. 293.

To the present generation, the *fact* of evolution is taken for granted. So it was for many students before Darwin. His contribution lay in the original suggestion of a plausible *method* of evolution. Part of this method involved an intense struggle for existence between organisms, resulting in the survival of the fittest. About this almost ceaseless warfare there was a fierce individualism. In Darwin's words, "every single organic being may be said to be striving to the utmost to increase in numbers; . . . each lives by a struggle at some period of its life" (*Origin of Species*). A universal rivalry and competition are today considered by some to be the sole "laws" of animal and human existence.

But may there be other methods of evolution in operation, perhaps factors of a cooperative nature? For over twenty years Professor Allee has been experimenting with animals to determine whether the emphasis on struggle in Darwin's hypothesis adequately described observed activities. He has come to the conclusion that "there is a general principle of automatic cooperation which is one of the fundamental biological principles" (50). In certain situations only those animals that work together tend to survive. Thus there is added an extremely important factor to our knowledge of the behaviors necessary for survival. The fittest are those who sometimes *cooperate* to help each other live and who sometimes *struggle* with each other for access to a limited food supply and optimum climatic conditions.

The idea of "mutual aid" has had a long yet not quite respectable scientific history, partly because of over-zealous backers who presented questionable evidence. Professor Allee and his students have been successful in collecting a mass of experimental data which demonstrates this less spectacular fact of cooperation for survival.

Most people know some of the harmful effects of overcrowding. There are less striking, though equally demonstrable, harmful effects from undercrowding. Goldfish, when grouped, live longer in toxic colloidal silver than when isolated; cold-blooded animals, when massed, withstand both higher and lower temperatures than when isolated; planaria survive exposure to ultra-violet radiation better grouped than isolated; isolated bacteria will not grow if placed in a culture containing gentian violet, whereas a colony will; sea-urchin eggs cleave more rapidly in a concentrated form than in a sparse population; various pro-

tozoa divide more rapidly when grouped, etc. The positive survival values of grouping in these cases are due to physiological changes. It may seem curious to label them cooperative. But Professor Allee feels that the most complicated expressions of social life are "based on phases of mass physiology and social biology which taken alone seem to be social by implication only" (275).

It has long been known that grouped animals do not behave in the same manner as isolated ones. Whether or not this difference has survival value is distinct from the fact that basic biological functions are affected by numbers. Although in almost all cases overcrowding retards growth, curiously enough undercrowding does likewise. Isolated white mice grow more slowly than their fellows grouped in twos and fours; the flour beetle reproduces most rapidly in populations of intermediate sizes; various kinds of fish eat more if grouped than if isolated. The results of grouping upon learning among lower animals are also observable. To be sure, isolated cockroaches learn a maze more rapidly and with fewer errors than their paired fellows; parrakeets also learn a maze more rapidly when isolated than when paired; yet goldfish learn to get food through a small door much more rapidly when grouped than when isolated. These are but a few of the examples which lead to the conclusion that throughout the animal kingdom, from protozoa to men, "there exists a broad substratum of partially social behavior. . . . From such a substratum, given suitable conditions, societies emerge now and again as they have among ants and men" (174).

Military commanders expect certain behaviors from enlisted men. Professors expect indications of respect and attention from their students. Human society is, in fact, based to a great extent upon these mutual expectations. It may seem surprising that there are equally predictable social behaviors among chickens. Upon their first meeting, two chickens either fight, or one retreats without fighting. In either case, the winner has a "peck-right." He can peck without being pecked in return. In small flocks there is a definite hierarchy. The top chicken pecks all those below him, the second chicken all except the top, and so on. But among pigeons social positions are not so rigid. There is no "peck-right"—no dominant-subordinate expectancy. The outcome of an encounter cannot be predicted, as it can with chickens. "Putting the matter somewhat facetiously, chickens appear to have developed the sort of 'line-organization' characteristic of a military system or a fascist state, while these pigeons, together with the ring doves, canaries and parrakeets, are more democratic" (189).

Without actual observation it is hazardous to apply this biological principle of cooperation directly to human affairs. It was equally dangerous to interpret human history solely in terms of struggle, without extended and critical experience. Yet Darwin's name has been used to justify almost all forms of regulated and unregulated fighting, since it was thought the unfit were thus eliminated and the race thereby improved. If struggle is present everywhere among animals, then modern war is inevitable, since man is a fighter. It is difficult to determine whether the belief in this doctrine may not have been a potent factor leading to war. What might result if people were convinced that human beings were "naturally" cooperative?

Concerning this problem of human struggle, Professor Allee asks three questions: "1. To what extent do the underlying biological relationships tend to bring about war? 2. Is war biologically justified by the results produced? 3. Can the basic principles of struggle and of co-operation work together in the international relations of men" (212)?

1. Modern war is not the simple expression of a biological struggle tendency. It is essentially a learned activity, quite possible to avoid. There are dominating humans, as well as ascendant chickens, but this fact does not necessarily lead to war. There are other methods by which their needs can be satisfied. The pressure of population against a limited food supply will not become great enough, at the present declining birth-rate, to make war inevitable. It is a question of adequate food distribution, rather than food shortage.

2. Instead of helping the fit survive, modern war helps the feebleminded, the diseased, and the incompetent. Only the fit are selected for active service, and their chances of living are greatly reduced. There is abundant evidence that famine and disease among the civilian population do not lead to stronger, healthier nations.

3. To avoid war, there is suggested an international federal union, composed of representatives of the various states, with a "relatively unbiased court of final judicial appeal, and certain potential police power" (236). The present League of Nations failed partly because the member states could withdraw at any time, and partly because there was no international police power to regulate the "peck-right" of certain states. In this biologically reasonable future, where would the struggle principle operate? "In our struggles with our physical environment, with disease germs and insects, we have ample opportunity for the struggle for existence" (242).

At this time, studies similar to *The Social Life of Animals* may have far-reaching results. The demonstration of a fundamental biological tendency for cooperation will certainly lend aid to those who are convinced that war is not necessary. This particular book is outstanding because Professor Allee has presented his experimental evidence as well as his theoretical discussions in a manner which should be a model: without cumbersome technicalities or a specialized vocabulary, with wit, and with the constant reminder that he does not have all the answers.

ARNOLD THOMSEN.

Syracuse University.

MENTAL CONFLICTS AND PERSONALITY. By Mandel Sherman. New York: Longmans, Green and Co., 1938. Pp. 319.

Presentations on motivation at greater length than the ordinary elementary textbook chapter are not many, especially those which are written with a full knowledge of general psychological concepts and psychological ways of thinking. The present volume is one of the few, however, and should prove to be especially useful to teachers who must include the topic of motivation in their work. It is not and does not pretend to be an original contribution to knowledge. Though heavily influenced by psychoanalytic contributions, it yet lacks most of the shoddy philosophy and bad psychology which ordinarily offend the trained psychologist in works written by those who know only psychoanalytic ways of speaking. An especially commendable feature is its presentation in terms of attitudes. Too often conflicts and repressions and the like are presented as though they appeared only in overt action. That they do not, any observer may discover, but here the psychology of attitudes so motivated is presented with unusually good judgment.

The style is often difficult in ways which make the reading of the book a rather dull experience. There are, it is true, good case examples, many of them, but here and there are places where examples are needed when they do not appear. Thus in places the book is heavily abstract. But, even where there

are these good examples of peculiar conduct, the style is still labored, lacking in clarity, and difficult to read. This factor will doubtless prove a handicap not easily overcome for beginning readers, although persons seeking to be well and correctly informed on the many different aspects and kinds of motivation problems will find here much of value, even though they must work a little to get it.

EDMUND S. CONKLIN.

Indiana University.

THE TROUBLED MIND—A GENERAL ACCOUNT OF THE HUMAN MIND, AND ITS DISORDERS AND THEIR REMEDIES. By Harry Roberts and Margaret Nelson Jackson. New York: E. P. Dutton & Co., Inc., 1939. Pp. 284.

According to its preface, this book is "an attempt to explain, in simple language, some of the outstanding conclusions of contemporary psychologists, with such criticism as the experience of the authors suggests." One of the authors "has been engaged in general medical practice for several decades; the other has for many years been a medical officer at a hospital devoted to the care of mental abnormals."

In their respective chapters, the authors appraise many current views, oppose extreme positions, and write clearly. The chapters fit together well. The authors discuss the human mind, conscious and unconscious; dynamic mechanisms, normal and abnormal; personality; the ages of life; mental disorders; mental defects; treatment; and legal safeguards for the mentally sick. They see good in structuralism, functionalism, behaviorism, reflexology, physiology, medicine, and psychoanalysis. For them, Freud has contributed much that is good in the field, also much that is rubbish.

Roberts acknowledges mechanical causation in all life, but assumes also non-mechanical causation, which he takes purpose to be. He says that man's innate impulses fall into three groups, according as the impulses further the individual, the community, or offspring especially. Thus submission and leadership come within the second group, while the coitive urge and tenderness, respectively, make up the third group. In this connection Roberts suspects that "half of the finest fruits of human civilization" spring not from the coitive but from the tenderness urge; yet on a later page he says that "all forms of true religious worship or ecstasy, as well as of artistic expression and artistic appreciation may be regarded as partial sex-energy sublimations."

Roberts holds that "the function of intellect is to guide, to sublimate, not to replace, inner promptings and vital aspirations"; that reasonable sublimation, along the lines of conditioning, is good; and that sublimation must be distinguished from repression, which effects dissociation. Complexes enter here, complexes in Hart's and Tansley's broad sense of the term. Habits likewise are important, together with innate differences in glandular and other organization. Dreams, he thinks, may involve all of these factors, also some racial memories. Repression results from conflict, of which he cites clear examples.

Jackson shows that, in the neuroses and psychoses and their treatment, organic and psychogenic factors are important, and that the whole individual is involved. Her discussion of psychotherapy, while broad and sensible on the whole, omits hypnoanalysis together with the actual, basic psychology of psychotherapy.

Jackson's descriptions of the various mental disorders make these chapters useful for reference or assignment for students of sociology and for students in psychology. The book as a whole is good for the general reader who

wants some current perspective rather than detailed criticism of concepts. Probably most psychologists would have liked to see the book more critical or analytical than it is of such concepts as submission, leadership, sublimation, complexes, dissociation, personality types, racial memories, the psychology of psychotherapy, and purpose.

W. S. TAYLOR.

Smith College.

COOPERATION AND COMPETITION *

The Arapesh eat little flesh;
They live secure, but futile.
They're not competitive or harsh,
As are the Kwakiutl.

The Eskimos are doughty foes.
Their wives they're always beating.
Bathonga buy their wives with hoes
And will not stand for cheating.

Bachiga think that food and drink
Should come from lone endeavor.
The Zuni, herding sheep in peace,
Cooperate forever.

Samoans feel the great ideal
Is helping one another.
Ojibwas try to stand alone,
And no one loves his brother.

The Maori loaned whate'er they owned
From Kingdom Come till now;
But interest rates are very high
Among the Ifugao.

The Iroquois, when but a boy,
Works hard for fame and prestige.
The Manus work for status, too;
Their clothes are but a vestige.

To steal and run is lots of fun
For every male Dakota.
Well, now we've studied every one
Down to the last iota.

(So, some cooperate, let's say,
While some hold competitions,
And some have freedom, in their way,
While some have prohibitions.)

The theme we treat is incomplete
In this most brief exposure.
We trust it will perseverate
Because it has no closure.

A. F. J.

* With apologies to Margaret Mead. Cf. Mead, M., *Cooperation and competition among primitive peoples*. New York: McGraw-Hill, 1937.

NEW BOOKS RECEIVED

BEAN, R. B. *The peopling of Virginia.* Boston: Chapman & Grimes, Inc., 1938. Pp. viii+302. Price $3.00.

BENDER, L. *A visual-motor Gestalt test and its clinical use.* New York: The American Orthopsychiatric Association, 1938. Pp. xi+176.

DOLLARD, J. *et al. Frustration and aggression.* New Haven: Yale University Press, 1939. Pp. viii+209. Price $2.00.

FARIS, R. E. L. AND DUNHAM, H. W. *Mental disorders in urban areas.* Chicago: University of Chicago Press, 1939. Pp. xxxviii+270. Price $2.50.

FLETCHER, P. *Life without fear.* New York: E. P. Dutton & Co., 1939. Pp. vii+111. Price $1.00.

GESELL, A. *et al. Biographies of child develpoment.* New York: Paul B. Hoeber, Inc., 1939. Pp. xvii+328. Price $3.75.

GORER, G. *Himalayan village.* London: Michael Joseph, Ltd., 1938. Pp. 510.

GUTHEIL, E. A. *The language of the dream.* New York: Macmillan Co., 1939. Pp. 286. Price $3.50.

HORNEY, K. *New ways in psychoanalysis.* New York: W. W. Norton & Co., 1939. Pp. 313. Price $3.00.

LANDIS, C. AND HUNT, W. A. *The startle pattern.* New York: Farrar & Rinehart, Inc., 1939. Pp. x+168. Price $2.50.

LUND, F. H. *Emotions, their psychological, physiological and educative implications.* New York: The Ronald Press, 1937. Pp. xiii+305. Price $3.50.

MIKESELL, W. H. *Mental hygiene.* New York: Prentice-Hall, Inc., 1939. Pp. xvi+456. Price $2.50.

PATTY, W. L. *A study of mechanism in education.* New York: Teachers College, Columbia University, 1938. Pp. v+183. Price $1.85.

POLLENS, B. *The sex criminal.* New York: The Macaulay Co., 1938. Pp. viii+211. Price $2.00.

SCHEIDEMANN, N. V. *Lecture demonstrations for general psychology.* Chicago: The University of Chicago Press, 1939. Pp. x+241. Price $2.50.

SIMPSON, R. H. *A study of those who influence and of those who are influenced in discussion.* New York: Teachers College, Columbia University, 1938. Pp. 89. Price $1.60.

SPENCER, D. *Fulcra of Conflict.* New York: World Book Co., 1939. Pp. xii+306. Price $2.40.

THORNTON, HENRY AND FREDA. *How to achieve sex happiness in marriage.* New York: The Vanguard Press, 1939. Pp. xii+155. Price $2.00.

WESTERHOF, A. C. *Representative psychologists.* Maryland: Pilot Publishing Co., 1938. Pp. vi+119.

WITTY, P. A., SKINNER, C. E., *et al. Mental hygiene in modern education.* New York: Farrar & Rinehart, Inc., 1939. Pp. x+539. Price $2.75.